AD_____

Award Number: W81XWH-07-1-0648

TITLE: Evaluating the Role of Genetic Markers in Prostate Cancer Progression: A Multiethnic Cohort Experience

PRINCIPAL INVESTIGATOR: Dr. Sara Strom

CONTRACTING ORGANIZATION: University of Texas M.D. Anderson Cancer Center
Houston, TX 77030

REPORT DATE: October 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-10-2011	Final	4 Sep 200 7- 3 Sep 2011

4. TITLE AND SUBTITLE
Evaluating the Role of Genetic Markers in Prostate Cancer Progression: A Multiethnic Cohort Experience

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-07-1-0648

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Sara Strom

E-Mail: yyamamur@mdanderson.rog

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas M.D. Anderson Cancer Center
Houston, TX 77030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Most prostate cancer(PCa) research has focused on risk, little is known about predictors of progression and even less about how these factors differ by ethnicity/race. There are strong racial disparities in mortality with African-Americans twice as likely to die from PCa comparedto Caucasians; very little data are available in Hispanics. Our goal was to identify markers of PCa progression in a multiethnic cohort(773 Caucasians, 361 African-Americans & 246 Mexican-Americans). Medical records for all participants have been abstracted, and we have updated vital status using the National Death Index. Using DNA extracted from archived specimens, we utilized a genome wide association study approach to identify polymorphisms associated with PCa progression; this approach has lead to the identification of different polymorphisms to be relevant in Caucasians vs. African-Americans. These results may offer insight into racial/ethnic differences in PCa progression and may be used to develop race-specific models for PCa progression.

15. SUBJECT TERMS
Prostate cancer, race, ethnicity, prognosis, polymorphisms

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	62	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION:

There is a paucity of information regarding markers/factors associated with prostate cancer (PCa) outcome in the United States, especially how these factors differ among racial/ethnic groups. African-American men are more likely to have poorer outcome relative to age and stage-matched Caucasian patients; and very little is known about prognosis and even less about factors that could predict progression among Hispanics. The overall goal of our research project is to identify molecular, epidemiological and clinical markers related to prostate cancer (PCa) progression in a multiethnic cohort of 1,380 PCa patients (773 Caucasians; 361 African Americans, and 246 Mexican Americans).

BODY:

Task 1 **Patient follow-up. (Months 1-30)**

a. Update patient follow-up data by checking clinical schedules and medical charts for updated information. Using a validated medical abstraction form, all patient charts will be abstracted.
b. Signed medical releases of information will be requested for care received outside of our institution. Copies of medical records will be requested.
c. Death certificates will be obtained for all participants identified as deceased.
d. Patients' self-reported recurrences (and subsequent treatments) and secondary cancers will be verified.
e. Data will be entered into existing databases.

Institutional patient records for all participant have been abstracted using the standardized form attached as Appendix A. In addition to baseline treatment information, we abstracted follow-up information, such as each prostate specific antigen (PSA) level and date, adjuvant care received, prostate-related care (including care related to complications following treatment (i.e., incontinence, impotence)), as well as additional cancer diagnoses. Institutional medical records were available electronically, and abstractions are performed using a paper form and were entered into an existing clinical database. Institutional patient records were matched by the institutional Tumor Registry to determine which of our study participants had a return visit to the institution within the past year, and the most recent visit was abstracted and the medical record abstraction was updated for each participant. All medical records were abstracted using the standardized form attached as Appendix A. The most recent clinical follow-up date at our institution was used as the "last date of contact" at the University of Texas MD Anderson Cancer Center (UTMDACC).

For patients for whom we do not have recent follow-up information at UTMDACC, we conducted telephone interviews to request follow-up information. We utilized several different options to obtain updated contact information for these individuals, including general internet searches, reverse address searches, and credit records. The Acxiom

Insight Collection service, which is an internet-based paid subscription database, was the most useful tool for us. Interviews to determine the current health of patients not returning to UTMDACC were conducted by telephone following a standardized protocol whereby individuals were called at least 5 times at different times of the day, as well as on weekends following the telephone script included as Appendix B. In addition, when the call attempts were not successful, we sent a letter to the patient at the last known valid address (with address correction requested) explaining that we are trying to contact them regarding their follow-up in a study and requesting that they contact us at their earliest convenience. Updated health and risk factor information was collected by trained interviewers, using a standardized questionnaire modified for this project (Appendix C). Following this methodology, we have an average of more than 10 years of follow-up for the patients included in this study.

After completing the Centers for Disease Control Institutional Review Board (IRB) application, we obtained approval to receive vital status from the National Death Index, as well as immediate and underlying causes of death for deceased individuals. We have linked our patient database to the NDI and updated vital status for all patients. In addition, we obtained IRB approval from the Texas Department of Health and Vital Statistics to link our patient database with the registered deaths in Texas and surrounding states. Date of death as well as cause of death when available have been recorded in the study database for all known decedents.

Task 2 **Evaluate Constitutional Markers of Genetic Susceptibility.** (Months 1-30)

a. Genotyping assays for all genes will be established, tested and validated by the Department of Epidemiology Genotyping Core (Months 1-24).

Genotyping has been completed in the Department of Epidemiology Genotyping Core using the Illumina Infinium II Assay.

b. Biological samples for all participants will be located and retrieved from study archive freezers (Months 1-3).

Using our laboratory tracking database, biological samples for this study were identified, located and retrieved from our freezer facility and transferred to the genotyping facility.

c. DNA will be extracted from banked specimens (Months 1-12).

DNA was extracted from all of the banked specimens. The DNA quality, quantity and purity were assessed for each sample to increase the likelihood of success of the genotyping. The extracted DNA was successfully used for the genotyping assays performed and reported below.

d. DNA samples will be plated for genotyping analyses – half the samples will be done in Year 2 and the other half will be done in Year 3 (Months 13 & 25)

All samples have been quantified, standardized, plated, and submitted for

genotyping.

e. Genotyping will be done for half the samples in Year 2 (Months 13-24) and the other half in Year 3 (Months 25-30).

Initial genotyping was done using the proposed methodology for 611 cases for MMP-1, 615 for e-cadherin, 433 for beta-2-adrenergic receptor, and 725 for cyclin D1. In our preliminary analyses, we have found significant differences with respect to genotypic frequency between racial/ethnic groups for MMP-1, beta-2-adrenergic receptor and cyclin D1. However, due to improvements in technology and published reports in recent literature, we have changed our genotyping methodology to utilize the Illumina platform for the final genotyping analyses. Using the Illumina Infinium II platform, we genotyped DNA samples from 1275 patients for 96 single nucleotide polymorphisms (SNPs) identified by genome wide association studies and validation studies to play a role in PCa risk. As a result of our collaborations with several multi-ethnic consortiums (led by Tim Rebbeck at University of Pennsylvania, Brian Henderson at the University of Southern California, and Ros Eeles at the Institute of Cancer Research Royal Cancer Hospital-London), several novel SNPs have been identified to play a role specifically in PCa risk among African-Americans (Appendix E).

Task 3 **Final Analysis and Preparation of Reports.** (Months 30-36)

The preliminary results of this study were presented in part at the 2011 Department of Defense IMPACT meeting. We have completed the final analyses of the data as proposed regarding the associations with disease progression and advanced stage at diagnosis, and are preparing to submit the manuscripts for consideration of publication. In addition, the results from the African-American population were included as part of our collaborations with the multi-ethnic consortiums (see Appendix E). We are finalizing the analyses and preparing manuscripts for submission regarding the associations with disease progression and advanced stage at diagnosis.

KEY RESEARCH ACCOMPLISHMENTS:

We found that different combinations of PCa susceptibility loci were associated with PCa outcome among Whites vs. AAs, and these loci differed between disease progression and metastatic at diagnosis. There were no significant differences between Hispanic and non-Hispanic Whites with respect to the associations with susceptibility loci. The data in concert with our previous work in PCa risk supports the finding that PCa in AAs may have a different etiologic basis vs. in Whites. Future research will focus on developing models to evaluate the role of these loci in determining subgroups of patients that may benefit from targeted early intervention to prevent disease progression/metastasis.

REPORTABLE OUTCOMES:

To date there have been 4 published papers (Appendix E), and there are 2 others currently pending review. There have been no patents or licenses applied for based on this award. Additionally, there have not been any degrees supported by this award; no cell lines, tissue or serum repositories developed; no informatics applied for based on work from this award; no employment opportunities applied for and/or received based on experience/training supported by this award. An abstract with these data was presented at the 2011 IMPaCT meeting. Preliminary data (numbers of participants with follow-up information) have been included in 2 funded grant proposals: U01-Genome-wide association study of prostate cancer in African Americans (Henderson); U19 –Trans-disciplinary cancer genomics research: post-GWA initiative (Henderson/Eeles).

CONCLUSION:

Our research may help explain ethnic/racial disparities in PCa progression and provide direction towards eliminating these disparities. Additionally, our results may guide future studies to develop ethnic/racial specific interventions (i.e., behavioral, clinical) to improve outcome in the most common cancer in American men.

REFERENCES:

Chang BL, Spangler E, Gallagher S, Haiman CA, Henderson B, Isaacs W, Benford ML, Kidd LR, Cooney K, Strom S, Ingles SA, Stern MC, Corral R, Joshi AD, Xu J, Giri VN, Rybicki B, Neslund-Dudas C, Kibel AS, Thompson IM, Leach RJ, Ostrander EA, Stanford JL, Witte J, Casey G, Eeles R, Hsing AW, Chanock S, Hu JJ, John EM, Park J, Stefflova K, Zeigler-Johnson C, Rebbeck TR. Validation of genome-wide prostate cancer associations in men of African descent. Cancer Epidemiol Biomarkers Prev. 2011 Jan; 20(10): 23-32.

Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, Hsing AW, Nemesure B, Rebbeck TR, Cooney KA, Xu J, Kibel AS, Hu JJ, John EM, Gueye SM, Watya S, Signorello LB, Hayes RB, Wang Z, Yeboah E, Tettey Y, Cai Q, Kolb S, Ostrander EA, Zeigler-Johnson C, Yamamura Y, Neslund-Dudas C, Haslag-Minoff J, Wu W, Thomas V, Allen GO, Murphy A, Chang BL, Zheng SL, Leske MC, Wu SY, Ray AM, Hennis AJ, Thun MJ, Carpten J, Casey G, Carter EN, Duarte ER, Xia LY, Sheng X, Wan P, Pooler LC, Cheng I, Monroe KR, Schumacher F, Le Marchand L, Kolonel LN, Chanock SJ, Berg DV, Stram DO, Henderson BE. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat Genet. 2011 Jun; 43(6):570-3.

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ, Bock CH, Boerwinkle E, Cai Q, Caporaso N, Casey G, Cupples LA, Deming SL, Diver WR, Divers J, Fornage M, Gillanders EM, Glessner J, Harris CC, Hu JJ, Ingles SA, Isaacs W, John EM, Kao WH, Keating B, Kittles RA, Kolonel LN, Larkin E, Le Marchand L, McNeill LH, Millikan RC, Murphy A, Musani S, Neslund-Dudas C, Nyante S, Papanicolaou GJ, Press MF, Psaty BM, Reiner AP, Rich SS, Rodriguez-Gil JL, Rotter JI, Rybicki BA, Schwartz AG, Signorello LB, Spitz M, Strom SS, Thun MJ, Tucker MA, Wang Z, Wiencke JK, Witte JS, Wrensch M, Wu X, Yamamura Y, Zanetti KA, Zheng W, Ziegler RG, Zhu X, Redline S, Hirschhorn JN, Henderson BE, Taylor HA Jr, Price AL, Hakonarson H, Chanock SJ, Haiman CA, Wilson JG, Reich D, Myers SR. The landscape of recombination in African Americans. Nature. 2011 Jul 20;476(7359):170-5.

Strom SS, Yamamura Y, Forman MR, Pettaway CA, Barrera SL, DiGiovanni J. Saturated fat intake predicts biochemical failure after prostatectomy. Int J Cancer. 2008 Jun 1;122(11):2581-5.

APPENDICES:

APPENDIX A:

Medical record abstraction form

Medical Records Abstraction Form

Name **MDACC# _____**

_____ MDA registration date _____/_____/____

Address

_____ **Date of birth** ____/_____/_____

_____ **Age at diagnosis _____years**

_____ **Phone number _____**

Ethnicity

☐ White ☐ Hispanic → ☐ African-American ☐ Asian ☐ Other _____	☐ Mexican ☐ Cuban ☐ S. American ☐ Other _____

Vital status ☐ Living ☐ Deceased → Date of death ____/____/____

Place of death _____

Cause of death _____

Last date of contact _____/_____/_____ Place of contact_____

Height: _____cm **Weight**: _____kg

_____ft/inches _____Ibs

Prostate cancer diagnosis

Date of diagnosis ____/_____/_____ Place of diagnosis: MDACC ☐ Yes ☐ No

Diagnostic tests ☐ Biopsy ☐ POS ☐ NEG ↓

☐ TURP ☐ POS ☐ NEG Where_____

☐ Chest x-ray ☐ POS ☐ NEG _____

☐ Bone scan ☐ POS ☐ NEG _____

☐ CT scan ☐ POS ☐ NEG

☐ Other _____☐ POS ☐ NEG When ____/_____/_____

Comments _____

_____

_____

Clinical stage of diagnosis

☐ Organ confined disease

☐ Regional disease

☐ Metastatic disease → date of confirmation ____/____/____

Sites: ☐ Bones ☐ Liver ☐ Adrenal gland ☐ Kidney ☐ Brain ☐ Other

_____

TNM stage

T1 → ☐ x ☐ 0 ☐ a ☐ b ☐ c **T2** → ☐ a ☐ b ☐ c **T3** → ☐ a ☐ b **T4**

N → ☐ x ☐ 0 ☐ 1 ☐ 2 ☐ 3

M → ☐ x ☐ 0 ☐ 1 Summary _____

Comments _____

_____

Laboratory results

Post-treatment values

Most recent post-treatment PSA value _____ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Follow-up PSA Values _____ ng/ml Date _____/_____/_____

Initial post-treatment PSA value _____ ng/ml Date _____/_____/_____

Pre-treatment values

Highest pre-treatment PSA value_____ ng/ml Date ____/_____/_____

Initial pre-treatment PSA value _____ ng/ml Date ____/_____/_____

Comments: _____
_____

Pathology report **Pathology report #:** _____

Specimen type ➲ ☐ **Prostatectomy**

MDACC grade ⇐ ☐ I ☐ II ☐ III ☐ IV ☐ other _____

Seminal Vesicle involvement ☐Yes ☐No **S/Margins** ☐ Positive ☐ Negative

Combined Gleason score

☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 _____

Dominant focus size /size _____cm Prostate volume _____ cm

Tumor locations ☐ Peripheral zone ☐ Central zone ☐ Transitional zone ☐ AFM zone

Comments_____
_____

Pathology report **Pathology report #:** _____

Specimen type ⇐ ☐ **Biopsy**

MDACC grade ⇐ ☐I ☐ II ☐ III ☐ IV ☐ other _____

Combined Gleason score

☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

Dominant focus size /size _____cm Prostate volume _____ cm

Tumor locations ☐ Peripheral zone ☐ Central zone ☐ Transitional zone ☐ AFM zone

Comments_____
_____

History of prostate cancer screening

☐ No

☐Yes → Type of screening test ☐ Prostate-specific antigen (PSA)

☐ Digital rectal examination (DRE)

☐ Trans-rectal ultrasound (TRUS)

☐ Other _____

Presence of urinary symptoms ☐ Yes ☐ No

Coments: _____
_____

Prostate cancer treatment received

☐ Radical prostatectomy Type → ☐ Radical Retropubic Prostatectomy (RRP) Date _____/_____/_____

☐ Radical perineal prostatectomy (RPP)

☐ Nerve-sparing

☐ Pelvic lymphadenectomy

☐ Orchiectomy → Date _____/_____/_____

☐ Cryosurgery → Date _____/_____/_____

		Onset of treatment	End of treatment
☐ Radiotherapy (EBRT)	→	Date _____/_____/_____	Date _____/____/____
☐ Brachytherapy	→	Date _____/_____/_____	Date _____/____/____
☐ Hormonal therapy	→	Date _____/_____/_____	Date _____/____/____
☐ Immunotherapy	→	Date _____/_____/_____	Date _____/____/____
☐ Surveillance	→	Date _____/_____/_____	Date _____/____/____
☐ Chemotherapy	→	Date _____/_____/_____	Date _____/____/____
☐ Other (specify) _____		Date _____/_____/_____	Date _____/____/____

Comments _____
_____

Complications of treatment

Urinary
Incontinence ☐ No
☐ Yes → Uses sanitary pad ☐ No
☐ Yes → number /day_____

Treatment received_____

Post-treatment status (1yr.) Number of pads/day _____ Date ____/____/____

Impotence ☐ No
☐ Yes → Treatment received _____
Post-treatment status (1yr.)_____

Urinary retention ☐ No
☐ Yes Treatment received
_____

Other_____
_____
_____

Comorbid conditions prior to diagnosis of prostate cancer

☐ No ☐ Yes

↓

☐ Diabetes (IDDM, NIDDM) Date of diagnosis ___/___/___

☐ Hemorrhage Date of diagnosis ___/___/___

☐ Hypertension Date of diagnosis ___/___/___
☐ Peptic ulcer disease Date of diagnosis ___/___/___

☐ Congestive heart failure Date of diagnosis ___/___/___
☐ Pancreatitis Date of diagnosis ___/___/___

☐ Myocardial infarction Date of diagnosis ___/___/___
☐ Cholelithiasis Date of diagnosis ___/___/___

☐ Stroke Date of diagnosis ___/___/___

☐ Alcoholism Date of diagnosis ___/___/___

☐ Chronic obstructive pulmonary disease Date of diagnosis ___/___/___
☐ Lupus erythematosus Date of diagnosis ___/___/___

☐ Other _____ Date of diagnosis ___/___/___

Other pertinent information

Recurrence of prostate cancer

☐ No
☐ Yes → Date of diagnosis _____/_____/_____

Place of diagnosis _____

Type of treatment _____

Basis of diagnosis _____
_____
_____

Diagnostic tests ☐ Biopsy ☐POS ☐NEG
☐ TURP ☐POS ☐NEG
☐ Chest x-ray ☐POS ☐NEG
☐ Bone scan ☐POS ☐NEG
☐ CT scan ☐POS ☐NEG
☐ Other _____☐POS ☐NEG

Conditions diagnosed after diagnosis of prostate cancer

☐ No ☐ Yes

↓

Date of diagnosis _____/_____/_____
Type of disease _____
Place of diagnosis _____
Type of treatment received_____

Date of diagnosis _____/_____/_____
Type of disease _____
Place of diagnosis _____
Type of treatment received_____

Comments _____
_____

Last clinic visit **Date _____/_____/_____**

Notes

APPENDIX B:

Follow-up telephone recruitment script

SCRIPT 1 (Speaking to person who answers phone) –

Hello, my name is (INTERVIEWER'S NAME) and I am calling on behalf of MD Anderson Cancer Center, here in Houston. May I please speak with (PATIENT'S NAME)?

- NOT AVAILABLE – Verify (PATIENT'S NAME) lives at this residence. Ask "Is there a time that I could call back and speak with him?" OR "would you please ask him to call me (INTERVIEWER'S NAME) at (PHONE NUMBER) at his earliest convenience? Thank you for your assistance.

- YES – Thank you…(Wait for (PATIENT'S NAME) come to phone) Hello, my name is (INTERVIEWER'S NAME) and I am calling on behalf of MD Anderson Cancer Center, here in Houston. You participated in one of our prostate cancer studies a few years ago, and we are conducting a follow-up study to see how you are doing. Would it be all right with you if I asked you a few questions about your health and updated your information?

  - NO – thank you for your time. If you change your mind and would like to participate, please contact me (INTERVIEWER'S NAME) at (PHONE NUMBER).

  - YES – I want to let you know that answering these questions is completely voluntary, and you may decide not to answer any or all of them. (Administer risk factor questionnaire (Appendix D))

Following each call, the interviewer logs each call made onto the tracking log for each file, documenting the date, time, phone number dialed, and with whom they spoke. These logs are maintained in the individual patient's study chart, kept in a locked office coded by study identification number.

APPENDIX C:

Follow-Up questionnaire

PROSTATE CANCER FOLLOW-UP STUDY

M.D. A

M.D. Anderson Cancer Center *Department of Epidemiology*

STUDY NUMBER:_____ **DATE OF PC DIAGNOSIS:** _____/_____/_____

MED RECORD/PATIENT #:_____

DATE OF BASELINE INTERVIEW: _____/_____/_____

PATIENT RECEIVING FOLLOW-UP CARE AT MDACC: __(1)YES __(2) NO

DATE OF MOST RECENT MDACC VISIT: ___/____/____

_____ **HOME PHONE:** (____)_____

FIRST NAME **M.I.** **LAST NAME**

_____ **WORK PHONE:** (____)_____

STREET ADDRESS

_____ **SSN:** _____

CITY **STATE** **ZIP CODE**

INTERVIEW DATE: _____/_____/_____ **INTERVIEWER'S INITIALS:** _____

WHO IS COMPLETING QUESTIONNAIRE? O **PATIENT** O **PROXY**

IF PATIENT IS DECEASED, DATE OF DEATH _____ **COUNTY & STATE OF DEATH** _____

As you may remember, you participated in a study of prostate cancer. We are currently updating our information, and we wanted to see how you are doing. Do you have a few moments to talk to me now or when can I call you back?

1. Are you currently being followed-up for your previous prostate cancer? _____ YES (1) _____ NO (2)

2. Where are/were you receiving follow-up care? _____

3. When was your most recent follow-up visit? _____ (Date)

When was the last time you had (the following test(s))? What were the results?

Test	Most Recent Date	Result (most recent)	
4. Prostate Specific Antigen/ (PSA)			___ Normal (1) **go to Q.8** ___ Abnormal (2) **go to Q.5**
5. Ultrasound (TRUS)			
6. Biopsy or Transurethral Resection of Prostate (TURP)			
7. Other (specify)			

8. Have you received any prostate treatment since you were last seen at MD Anderson/Kelsey-Seyboldt/VAMC/Dr. _____ (select provider) in _____ (fill in last date)?

____ (1)YES _____(2) NO → Skip to Q. 12

9. When and where were/are you receiving treatment? (*e.g., MD/Clinic Name, Address, Phone #*)

Office Note: Obtain signed medical release of information

_____

_____

10. What type(s) of treatment did you receive? (e.g., radiation, hormone shots, hormone pills, chemotherapy)

_____

_____

11. Why was the treatment necessary?

_____

_____

Have you ever been told by a doctor or another health care professional that you have any of the following conditions?

CONDITION	BEEN TOLD?	DATE/AGE DIAGNOSED	TREATMENT/MEDICATION NAME
12. Diabetes (or sugar in urine)	_____ (1) YES _____ (2) NO		
13. Hypertension (high blood pressure)	_____ (1) YES _____ (2) NO		
14. Angina (angina pectoris)	_____ (1) YES _____ (2) NO		

15. Heart attack (myocardial infarction)	_____ (1) YES _____ (2) NO		
16. Any other kind of heart condition or disease (not mentioned above) SPECIFY: _____	_____ (1) YES _____ (2) NO		
CONDITION	BEEN TOLD?	DATE/AGE DIAGNOSED	TREATMENT/MEDICATION NAME
17. High cholesterol	_____ (1) YES _____ (2) NO		
18. Arthritis TYPE: _____	_____ (1) YES _____ (2) NO		
19. Any other cancer(s)? SPECIFY	_____ (1) YES _____ (2) NO		
20. Any other condition(s)? SPECIFY	_____ (1) YES _____ (2) NO		

TOBACCO

The next questions are about smoking.

Previous Smoking Status
____ Current ____ Former ____ Never

21. Since your prostate cancer diagnosis, has your smoking status changed? ___(1)YES ___(2) NO → **Fmr/Never smoker** Go to Q.24 **Currt smkr** Go to Q.23

22. Are you currently smoking cigarettes? ____ (1)YES ____(2) NO ⟶ When did you stop? _____ (Year)

23. On average, how many cigarettes per day do you/did you smoke? _____

MEDICATION/SUPPLEMENT USE

The next questions are medications and supplement use

24. Have you taken any supplements, over the counter medications or prescription medications at least once a month since your diagnosis? This would include all vitamins, minerals, herbal and non-herbal supplements of any kind.

_____ (2) No, GO TO Q. 26

_____ (1) Yes, Fairly regularly _____ (3)Yes, but NOT regularly

25. Please list the names of any supplements (including vitamins, minerals and herbal supplements), over-the-counter medications or prescription medications that you have taken. Also include the number of pills or tablets taken daily, weekly, monthly or yearly?

For Office Use:	_____ code

Supplement, Over-the-counter or prescription medication	**Number per Day**	**Number per Week**	**Number per Month**	**Number per Year**	**Rarely / Never (✓)**	**How many years?**	**Dose**
Brand:_____ Name on bottle: _____							
Brand:_____ Name on bottle: _____							
Brand:_____ Name on bottle: _____							

Brand:_____ Name on bottle: _____							

DIET

The following questions are regarding diet changes

Since your diagnosis, have you changed your consumption of the following types of foods?

FOOD TYPE	INCREASED
26. Fat	____(1) increased ____(2) decreased ____(3) no change
27. Fruits	____(1) increased ____(2) decreased ____(3) no change
28. Vegetables	____(1) increased ____(2) decreased ____(3) no change
29. Fiber	____(1) increased ____(2) decreased ____(3) no change
30. Soy products	____(1) increased ____(2) decreased ____(3) no change

31. Are there any comments that you would like to add about your diet or about the way you have changed your diet?

_____

_____

FAMILY HISTORY

In this section, I would like to ask you some questions about your family

FAMILY HISTORY PRE-CODE:

Previously reported family members WITH cancer:

Sex	Relative	Side of Family	Type of Cancer	Sex	Relative	Side of Family	Type of Cancer

32. Previously, you told us that your _____ (insert previous history here) had cancer, have any other immediate family members been diagnosed with cancer? ____ YES (1) ____ NO (2) → Go to Q. 34

33. Would you please give us some information about these NEW family members diagnosed with cancer? (DON'T include those previously reported)

Rel Code	Sex	Relative	Rel UIN	When was he/she born?	What kind of cancer? ICD-9		When was he/ she diagnosed?	Is he/she still living?	When did he/she die?
								____(1) Yes ____(2) No	
								____(1) Yes ____(2) No	

								____(1) Yes ____(2) No	
								____(1) Yes ____(2) No	
								____(1) Yes ____(2) No	

OCCUPATIONAL HISTORY

In this section, I would like to ask you some questions about your current occupation

34.

What is your job or occupation?	Years employed	Major duties	Equipment used (Any Chemicals?)	Work done by company	SIC	OCC
Current Job:	_____**To**_____					
Spec						

If we need additional information from you in the future, can we contact you by telephone? ____ (1)YES ____ (2)NO

This is the end of our interview. I would like to thank you for your help with our research. If you have any questions that I or Dr. Strom can answer in the future, please feel free to contact us. We would also like to verify that we have your current address correctly recorded. We have your current address as: ***READ ADDRESS FROM FILE RECORD***

Is this address correct? _____(1) YES _____ (2)NO *(If NO, please provide correct information below)*

_____ _____ _____
First Name Middle Name Last Name

_____
Street Address

_____ _____ _____
City State Zip Code

Also, so that we may keep contact with you, would you please give me that name, address, and telephone number of a person who does not live with you who will know your whereabouts in the future:

_____ _____ _____
First Name Middle Name Last Name

_____
Street Address

_____ _____ _____
City State Zip Code

Thank you once again for your time and help with our research project. If we have any more questions in the future, we hope we can call you again.

INTERVIEW ASSESSMENT

Date of interview: ____/____/____ Interviewer's Initials: _____

Time Interview began: _____ Time Interview ended: _____

1. Respondent's cooperation was:
   - _____ Very Good (1)
   - _____ Good (2)
   - _____ Fair (3)
   - _____ Poor (4)

2. The quality of the interview was:
   - _____ Highly Reliable (1)
   - _____ Generally Reliable (2)
   - _____ Questionable (3)
   - _____ Unsatisfactory (4)

Please write comments about the interview: _____

_____

_____

_____

APPENDIX D:

Medical release of information form

AUTHORIZATION FOR DISCLOSURE OF HEALTH INFORMATION

(1) I hereby authorize _____ to disclose the following information from the health records of:

Patient Name: _____
Last First MI. Date of Birth

MDA #

Address:_____
_____
Street City State Zip Code

Phone

covering the period of healthcare from _____ to _____.

(2) Information to be disclosed:

- ❑ Complete Health Record
- ❑ Primary Medical Evaluation
- ❑ Progress Notes
- ❑ X-Ray Reports
- ❑ Discharge Summary
- ❑ Consultation Reports
- ❑ Laboratory Tests
- ❑ Radiotherapy Notes
- ❑ Chemotherapy Notes
- ❑ Nurse's Notes

❑ Other (specify) _____

I understand that this will include information relating to (check if applicable):

- ❑ Acquired Immunodeficiency Syndrome (AIDS) or infection with HIV (Human Immunodeficiency Virus)
- ❑ Psychiatric care
- ❑ Treatment for alcohol and/or drug abuse

(3) This information is to be disclosed to: Dr. Sara Strom

_____

Investigator's signature

UT MD Anderson Cancer Center

1515 Holcombe, Houston, Texas 77030

for the purpose of: **Medical Record completion for research protocol M91-004**.

(4) I understand this authorization may be revoked in writing at any time, except to the extent that action has been taken in reliance on this authorization. Unless otherwise evoked, this authorization will expire on the following date, event, or condition:

_____

(5) The facility, its employees, officers, and physicians are hereby released from any legal responsibility or liability for disclosure of the above information to the extent indicated and authorized herein.

Signed: _____
(patient) (date)

or _____
(Legal Representative)(Relationship to Patient) (date)

SUPPORTING DATA: N/A

APPENDIX E:

Published manuscripts

Chang BL, Spangler E, Gallagher S, Haiman CA, Henderson B, Isaacs W, Benford ML, Kidd LR, Cooney K, Strom S, Ingles SA, Stern MC, Corral R, Joshi AD, Xu J, Giri VN, Rybicki B, Neslund-Dudas C, Kibel AS, Thompson IM, Leach RJ, Ostrander EA, Stanford JL, Witte J, Casey G, Eeles R, Hsing AW, Chanock S, Hu JJ, John EM, Park J, Stefflova K, Zeigler-Johnson C, Rebbeck TR. Validation of genome-wide prostate cancer associations in men of African descent. Cancer Epidemiol Biomarkers Prev. 2011 Jan; 20(10): 23-32.

Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, Rybicki BA, Isaacs WB, Ingles SA, Stanford JL, Diver WR, Witte JS, Hsing AW, Nemesure B, Rebbeck TR, Cooney KA, Xu J, Kibel AS, Hu JJ, John EM, Gueye SM, Watya S, Signorello LB, Hayes RB, Wang Z, Yeboah E, Tettey Y, Cai Q, Kolb S, Ostrander EA, Zeigler-Johnson C, Yamamura Y, Neslund-Dudas C, Haslag-Minoff J, Wu W, Thomas V, Allen GO, Murphy A, Chang BL, Zheng SL, Leske MC, Wu SY, Ray AM, Hennis AJ, Thun MJ, Carpten J, Casey G, Carter EN, Duarte ER, Xia LY, Sheng X, Wan P, Pooler LC, Cheng I, Monroe KR, Schumacher F, Le Marchand L, Kolonel LN, Chanock SJ, Berg DV, Stram DO, Henderson BE. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat Genet. 2011 Jun; 43(6):570-3.

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ, Bock CH, Boerwinkle E, Cai Q, Caporaso N, Casey G, Cupples LA, Deming SL, Diver WR, Divers J, Fornage M, Gillanders EM, Glessner J, Harris CC, Hu JJ, Ingles SA, Isaacs W, John EM, Kao WH, Keating B, Kittles RA, Kolonel LN, Larkin E, Le Marchand L, McNeill LH, Millikan RC, Murphy A, Musani S, Neslund-Dudas C, Nyante S, Papanicolaou GJ, Press MF, Psaty BM, Reiner AP, Rich SS, Rodriguez-Gil JL, Rotter JI, Rybicki BA, Schwartz AG, Signorello LB, Spitz M, Strom SS, Thun MJ, Tucker MA, Wang Z, Wiencke JK, Witte JS, Wrensch M, Wu X, Yamamura Y, Zanetti KA, Zheng W, Ziegler RG, Zhu X, Redline S, Hirschhorn JN, Henderson BE, Taylor HA Jr, Price AL, Hakonarson H, Chanock SJ, Haiman CA, Wilson JG, Reich D, Myers SR. The landscape of recombination in African Americans. Nature. 2011 Jul 20;476(7359):170-5.

Strom SS, Yamamura Y, Forman MR, Pettaway CA, Barrera SL, DiGiovanni J. Saturated fat intake predicts biochemical failure after prostatectomy. Int J Cancer. 2008 Jun 1;122(11):2581-5.

Cancer Epidemiology, Biomarkers & Prevention

Validation of Genome-Wide Prostate Cancer Associations in Men of African Descent

Bao-Li Chang, Elaine Spangler, Stephen Gallagher, et al.

Cancer Epidemiol Biomarkers Prev 2011;20:23-32. Published OnlineFirst November 11, 2010.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

Cancer Epidemiology, Biomarkers & Prevention

Research Article

Validation of Genome-Wide Prostate Cancer Associations in Men of African Descent

Bao-Li Chang[1], Elaine Spangler[1], Stephen Gallagher[1], Christopher A. Haiman[2], Brian Henderson[2], William Isaacs[3], Marnita L. Benford[4], LaCreis R. Kidd[4], Kathleen Cooney[5], Sara Strom[6], Sue Ann Ingles[7], Mariana C. Stern[7], Roman Corral[7], Amit D. Joshi[7], Jianfeng Xu[8], Veda N. Giri[9], Benjamin Rybicki[10], Christine Neslund-Dudas[10], Adam S. Kibel[11], Ian M. Thompson[12], Robin J. Leach[12], Elaine A. Ostrander[13], Janet L. Stanford[14], John Witte[15], Graham Casey[16], Rosalind Eeles[17], Ann W. Hsing[18], Stephen Chanock[18], Jennifer J. Hu[19], Esther M. John[20], Jong Park[21], Klara Stefflova[1], Charnita Zeigler-Johnson[1], and Timothy R. Rebbeck[1,22]

Abstract

Background: Genome-wide association studies (GWAS) have identified numerous prostate cancer susceptibility alleles, but these loci have been identified primarily in men of European descent. There is limited information about the role of these loci in men of African descent.

Methods: We identified 7,788 prostate cancer cases and controls with genotype data for 47 GWAS-identified loci.

Results: We identified significant associations for SNP rs10486567 at *JAZF1*, rs10993994 at *MSMB*, rs12418451 and rs7931342 at 11q13, and rs5945572 and rs5945619 at *NUDT10/11*. These associations were in the same direction and of similar magnitude as those reported in men of European descent. Significance was attained at all reported prostate cancer susceptibility regions at chromosome *8q24*, including associations reaching genome-wide significance in region 2.

Conclusion: We have validated in men of African descent the associations at some, but not all, prostate cancer susceptibility loci originally identified in European descent populations. This may be due to the heterogeneity in genetic etiology or in the pattern of genetic variation across populations.

Impact: The genetic etiology of prostate cancer in men of African descent differs from that of men of European descent. *Cancer Epidemiol Biomarkers Prev; 20(1) ; 23–32. ©2011 AACR.*

Introduction

The differences in prostate cancer incidence and mortality across men of different racial groups are well documented. According to SEER, prostate cancer has an age-adjusted incidence rate of 234.6 per 100,000 in African American and 150.4 per 100,000 in European American men. In addition, a 2.4-fold difference in mortality rate (62.3 per 100,000 in African Americans vs. 25.6 per 100,000 in European Americans) represents the greatest disparity between these groups of any major cancer site. Despite this profound public health concern, knowledge of the etiologic underpinnings for this disparity remains unclear. It is likely that inherited susceptibility,

Authors' Affiliations: [1]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; [2]Department of Preventive Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California; [3]The Johns Hopkins University School of Medicine and Bloomberg School of Public Health, Baltimore, Maryland; [4]University of Louisville, Louisville, Kentucky; [5]Department of Medicine, University of Michigan Medical School, Ann Arbor, Michigan; [6]MD Anderson Cancer Center, Houston, Texas; [7]Department of Preventive Medicine, University of Southern California, Keck School of Medicine, Los Angeles, California; [8]Wake Forest University, Winston-Salem, North Carolina; [9]Fox Chase Cancer Center, Philadelphia, Pennsylvania; [10]Henry Ford Hospital, Detroit, Michigan; [11]Department of Surgery and Genetics, Washington University School of Medicine, St. Louis, Missouri; [12]Department of Urology and the Cancer Therapy and Research Center, University of Texas Health Science Center at San Antonio, Texas; [13]National Human Genome Research Institute, Bethesda, Maryland; [14]Fred Hutchinson Cancer Research Center, Seattle, Washington; [15]Departments of Epidemiology & Biostatistics, Urology, Institute for Human Genetics, University of California, San Francisco, California; [16]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; [17]Institute for Cancer Research and Royal Marsden NHS Trust, Sutton, United Kingdom; [18]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; [19]University of Miami Miller School of Medicine, Miami, Florida; [20]Northern California Cancer Center, Union City, California; [21]Moffitt Cancer Center, Tampa, Florida; and [22]Abramson Cancer Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers and Prevention Online (http://cebp.aacrjournals.org/).

R. Eeles: for The UKGPCS Coordinating Group and Collaborators.

Corresponding Author: Timothy R. Rebbeck, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 217 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104. Phone: 215-898-1793; Fax: 215-573-1050. E-mail: rebbeck@mail.med.upenn.edu

doi: 10.1158/1055-9965.EPI-10-0698

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

environmental exposures, lifestyle, behavior, screening, and cancer treatment all influence the disparity between men of different racial and ethnic backgrounds.

A number of recent genome-wide association studies (GWAS) have identified numerous prostate cancer susceptibility loci including *CTBP2* (chr. 10q26), *EHPB1* (chr. 2p15), *HNF1B* (chr. 17q12), *IGF2/IGF2A/INS* (chr. 11p15), *ITGA6* (chr. 2p31), *KLK2/3* (chr. 19q13), *LMTK2* (chr. 7q21), *MSMB* (chr. 10q11), *NKX3.1* (chr. 8p21), *NUDT10/11* (chr. Xp11.22), *PDLIM5* (chr. 4q22), *SELB* (chr. 3q21.3), *SLC22A3* (chr. 6q25), *TET2* (chr. 4q24), *THADA* (chr. 2p21), *TTLL1/BIK/MCAT/PACSIN2* (chr. 22q13), as well as loci on chromosome 11q13, 17q12, 17q24, and multiple regions at chromosome 8q24 (1–17). These loci were discovered primarily in European descent men (EDM), with the exception being the prostate cancer susceptibility loci at chromosome 8q24, which were identified by linkage and admixture mapping (15, 18). Studies suggest that some genetic variants confer risk across populations but with different magnitudes of the risk in different populations, or they may only confer risk in one population but not in others (11, 19). Because the prevalence of prostate cancer and the allele frequencies differ between EDM and African descent men (ADM), it is important to estimate the effects of these GWAS risk variants originally identified in EDM on ADM before generalization of the GWAS associations in ADM. Three recent studies have also attempted to validate associations between some of the loci listed above and prostate cancer in ADM. Xu et al. (20) studied 868 cases and 878 controls and validated the loci at 8q24 ($P = 0.034 - P = 2 \times 10^{-5}$) and 3p12 ($P = 0.029$). Waters et al. (19) studied 860 cases and 575 controls and validated *KLK2/3* (19q13.33) and *NUDT10/11* (Xp11.22). Finally, Hooker et al. (21) validated 8q24 ($P = 1 \times 10^{-4}$), 11q13.2 ($P = 0.009$), *HNF1B/TCF2* (17q12; $P = 0.008$), *KLK2/3* (19q13.33; $P = 0.04$), and *NUDT11* (Xp11.22; $P = 0.05$) in 454 cases and 301 controls. The validated loci were not consistent across these studies, perhaps due to relatively small sample sizes in each study. To confirm associations at previously identified prostate cancer susceptibility loci in ADM, we obtained data from 7,788 ADM from 19 centers in the United States and the United Kingdom for pooled analyses of GWAS-identified loci and prostate cancer.

Methods

Study sample

The sample studied here consisted of 4,040 cases and 3,748 controls ascertained from 19 centers (Supplementary Table 1). A detailed description of each center's study is presented in Appendix 1 and a summary of the study methods is presented in Supplementary Table 5. These studies include the Prostate Cancer Genetics Studies (CaP Genes) at the University of California (22), Fred Hutchinson Cancer Research Center (FHCRC) Prostate Cancer Studies (23, 24), The Prostate Risk Assessment Program (PRAP) at Fox Chase Cancer Center (25), The Flint Men's Health Study (FMHS; refs. 26, 27), Gene-Environment Interaction in Prostate Cancer (GECAP) Study at Henry Ford Hospital (28), Los Angeles County Study (LACS; ref. 29), Prostate Cancer Clinical Outcome Study (PC^2OS) at the University of Louisville (30), MD Anderson Cancer Center (31), The Multiethnic Cohort Study (MEC; ref. 32), Moffitt Cancer Center Study (33), NCI Prostate Tissue Study (NCIPTS), University of Pennsylvania Study of Cancer Outcomes, Risk, and Ethnicity (SCORE; ref. 34), University of Texas San Antonio Center for Biomarkers of Risk for Prostate Cancer (SABOR), University of Texas Health Science Center at San Antonio (35, 36), San Francisco Bay Area Prostate Cancer Study (SFBAPCS; ref. 37), United Kingdom Genetic Prostate Cancer Study (UKGPCS), Wake University Consortium including participants from the Johns Hopkins University, Wake Forest University, and Washington University (20). Two of these studies, SFBAPCS and UKGPCS, have contributed only to case–case analyses of disease aggressiveness because only cases were available from these 2 studies. Single-nucleotide polymorphism (SNP) were chosen if they were implicated in previous GWAS studies (1–3, 38), in follow-up fine-mapping studies (5–7, 39, 40), or associated with disease aggressiveness (4, 41). Available SNPs in all regions of 8q24, some of which were initially identified through linkage and admixture mapping in ADM and confirmed in GWAS studies, were also included (10, 11, 14–16, 42).

Genotype data were excluded if they were found to have genotyping failure rates greater than 5% within each study center or if they deviated significantly from Hardy–Weinberg proportions. We set a threshold of $P < 0.001$ based on multiple-test adjustment for the number of SNPs tested (family-wise error rate $P = 0.05$ divided by 50 SNPs equals to $P = 0.001$). SNPs were included in the present analysis if we obtained at least 1,000 genotypes in cases and controls from the contributing centers by October 2009. A summary of the data contributed by each center by SNP is summarized in Supplementary Table 6.

Statistical methods

Departure from Hardy–Weinberg equilibrium was assessed for each SNP in control subjects of the combined study populations using the chi-square goodness-of-fit Test. Any SNP that showed departure from Hardy–Weinberg equilibrium with $P < 0.001$ in controls was excluded from subsequent analyses. Unconditional logistic regression models were used to estimate odds ratios (OR) and 95% CIs to measure the association between individual SNP genotypes and prostate cancer risk or disease aggressiveness defined as Gleason score <7 versus 7+ or tumor stage T1/T2 versus T3/T4. Analyses were undertaken using an additive mode of inheritance, adjusting for age and study centers (results shown in Table 1 and Supplementary Tables 2–4).

Subgroup analyses were also carried out to estimate whether African ancestry affected the reported

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

associations. This analysis included a subset of study centers for which estimated percentage of African ancestry was available (Supplemental Table 5). Centers used different ancestry informative marker (AIM) panels (Supplementary Table 5). These AIMs were obtained from the original genotyping methods used in each center, and were comparable on the basis of several measures of marker informativeness (FST, FIC, and δ). The statistical methods used to estimate ancestry proportion, STRUCTURE and ANCESTRYMAP, have used same hierarchical model and probabilistic measures and would result in similar/high correlated measurements. In addition, we analyzed data stratifying by center to adjust for potential confounding by ancestry proportion within each participating study and to minimize the influence of varying informativeness of AIM panels.

These studies include nested case–control studies from within cohorts, matched and unmatched case–control studies, as well as case-only series. To address the potential study heterogeneity, age-adjusted ORs and 95% CIs for SNPs were estimated for each study population separately, and forest plots were generated for independent SNPs with P values < 0.05 (Supplementary Fig. 1). Potential heterogeneity in the association of SNPs with prostate cancer among study populations was examined by Breslow–Day homogeneity test. All statistical analyses were performed using SAS 9.2 and PLINK (43). An LD heat map (Fig. 1) was generated on the basis of HapMap YRI data using Haploview (44). Inferences were made using 2-sided hypothesis testing with a P value < 0.05. Because this is a validation study, we did not correct for multiple hypothesis tests.

Results

We were able to validate some, but not all, prostate cancer GWAS loci (Table 1 for SNPs outside of 8q24 regions, and Supplementary Table 2 and Figure 1 for SNPs located within 8q24). Most associations reported here were in the same direction and with an equal or smaller magnitude as those originally reported in EDM. However, a number of associations reported here were not in the same direction as those reported in EDM (i.e., CTBP2, 11q13, and 22q13; Table 1), suggesting that these alleles are not consistent with prostate cancer risk in ADM. A number of loci that were implicated in EDM were not associated with prostate cancer risk in ADM. These included *CTBP2* (rs4962416), 11q13 (rs12418451), *IL16* (rs4072111), *CDH13* (rs4782726), and 22q13 (rs9623117) with OR < 1 (i.e., in the opposite direction from that reported in EDM), and *EHBP1* (rs721048), *LMTK2* (rs6465657), *MINPP1* (rs12771728), Chromosome 12 (rs902774), and *KLK2/3* (rs887391) with OR near 1.0. Furthermore, the upper bound of the 95% CI for a number of loci in ADM did not overlap at least earlier estimates made in EDM, including 3p12.1 (rs2660753), *DAB2IP* (rs1571801), *MSMB* (rs10993994), *CTBP2* (rs4962416), *HNF1B* (rs4430796 and rs7501939), *KLK2/3* (rs2735839), 22q13 (rs9623117), and *NUDT10/11* (rs5945572 and rs5945619). These results suggest that some loci with genome-wide significance in non-African descent populations may not be associated with prostate cancer or may not have the same magnitude of effect in ADM.

Several SNPs showed statistically significant associations. SNPs in *JAZF1* (rs10486567; OR = 1.18; $P = 0.0002$), *MSMB* (rs10993994; OR = 1.12; $P = 0.005$), 11q13 (rs10896449; OR = 1.12; $P = 0.031$ and rs7931342; OR = 1.15; $P = 0.014$) and *NUDT10/11* (rs5945572; OR = 1.11; $P = 0.02$ and rs5945619; OR = 1.09; $P = 0.039$) were statistically significantly associated with prostate cancer risk. The direction of effect of each of these associations was in the same direction as those reported in EDM (Table 1).

We also undertook a similar analysis that excluded data that have been published previously to isolate a subset of study centers for evaluating further evidence of independent replications (19, 20). After excluding data from those studies (i.e., JHU, MEC, Wake-Hu, Wake-NC, and Wash U), both *JAZF1* rs10486567 ($P = 0.005$) and *MSMB* rs10993994 ($P = 0.009$) remained statistically significant. In both cases, the OR estimate in the subset trended away from the null hypothesis (OR = 1.23 in the subset vs. 1.18 in the total sample, and OR = 1.17 in the subset vs. 1.12 for the total sample, respectively). SNP rs10896449 at 11q13 stayed nominally significant ($P = 0.02$), but SNPs at *NUDT10/11* and SNP rs7931342 at 11q13 were no longer significant. These results further provide support for the association of *JAZF1* and *MSMB* with prostate cancer risk in ADM. Although we were unable to mutually adjust for the effects of multiple SNPs in a single locus for the majority of loci, after mutually adjusting for multiple SNPs at 11q13, both SNP rs7931342 (OR = 1.0; 94% CI: 0.77–1.30; $P = 0.999$) and rs10896449 (OR = 1.18; 95% CI: 0.93–1.49; $P = 0.17$) became nonsignificant. As the sample size for this last analysis is smaller than for the overall sample (i.e., $n = 2{,}013$ vs. $n = 3{,}954$ or $4{,}463$), we were not able to unambiguously determine which SNP contributed independently to the association signal seen at this locus. After mutual adjustment, the point estimates for rs7931342 changed from 1.15 to 1.0 and rs10896449 changed from 1.12 to 1.18. These results suggest that rs10896449 or other SNPs in tight LD with rs10896449 maybe the SNP that contributes to the association signal at 11q13 locus. Multiple independent loci on chromosome 8q24 have been identified as playing a role in prostate cancer etiology. We were able to validate the association of each of these regions at 8q24 (Fig. 1 and Supplementary Table 2). We had statistically significant evidence at the genome-wide association level for associations with regions 2 (rs13254738, rs6983561, and rs16901979), and statistically significant associations in region 1 (rs10090154), region 3 (rs6983267 and rs7000448), region 4 (rs7008482), and the region centromeric to region 2 (rs10086908). We also removed data that had been included in previous studies (19, 20, 45) of loci at 8q24. Significant associations remained for Regions 2 (block 2), 3 (block 4), 4 (centromeric to block 1), and the region

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

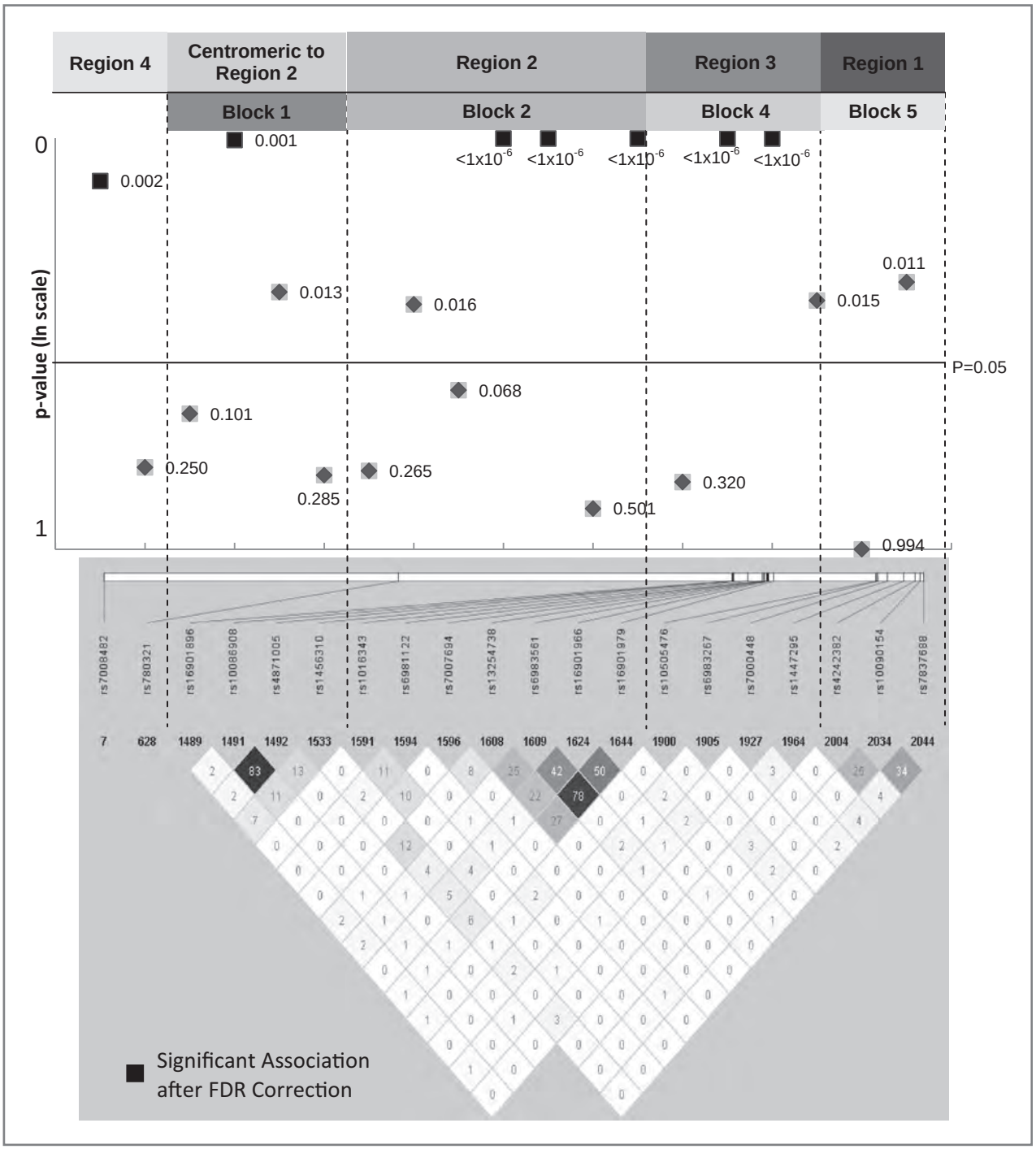


Figure 1. Results of prostate cancer associations at 8q24 in ADM. *P* values for association by genomic location.

centromeric to Region 2 (block 1). However, the marginal associations in region 1 (block 5) were no longer significant after the data from the published reports were excluded.

Because we have studied an admixed population of ADM, we also investigated potential bias due to population stratification by comparing the association results with or without adjusting for percentage of non-African ancestry estimated from AIMs. Ancestry adjustment analyses were undertaken in 8 of the 19 centers for which AIMs data were available (Supplementary Table 3). We observed significant differences in the proportion of African ancestry across centers (χ^2_7, Kruskal–Wallis = 339.6; $P < 0.0001$). However, these differences may reflect

Table 1. Results of associations at prostate cancer GWAS loci in men of African descent

Chr	Locus	SNP	Position, bp	Risk allele	Sample size		Risk allele frequency		aOR (95% CI)[a]	*P* value	OR in EDM[b]	RAF in EDM[b]
					Cases	Controls	Cases	Controls				
2	*EHBP1*	rs721048	62,985,234	A	2,083	2,116	0.05	0.04	1.02 (0.82–1.26)	0.882	1.15 (1)	0.19
3	*3p12.1*	rs2660753	87,193,363	T	1,796	1,415	0.50	0.49	1.05 (0.95–1.17)	0.338	1.18 (2)	0.11
6	*SLC22A3*	rs9364554	160,753,653	T	2,652	2,659	0.07	0.07	1.05 (0.90–1.24)	0.513	1.17 (2)	0.29
7	*JAZF1*	rs10486567	27,943,087	C	2,899	2,538	0.74	0.71	1.18 (1.08–1.29)	0.0002	1.21 (3)	0.77
7	*LMTK2*	rs6465657	97,654,262	C	3,266	2,631	0.87	0.87	1.03 (0.91–1.15)	0.666	1.12 (2)	0.46
8	*8p21*	rs6982080	29,686,860	G	844	536	0.28	0.28	1.08 (0.90–1.29)	0.432	1.03 (3)	0.31
9	*DAB2IP*	rs1571801	123,467,193	T	1,709	1,413	0.15	0.14	1.07 (0.92–1.24)	0.385	1.30 (4)	0.24
10	*MSMB*	rs7920517	51,202,626	G	849	541	0.74	0.71	1.13 (0.93–1.36)	0.217	1.22 (2)	0.48
10	*MSMB*	rs10993994	51,219,501	T	3,374	2,982	0.62	0.60	1.12 (1.03–1.21)	0.005	1.25 (2)	0.4
10	*MSMB*	rs7904463	51,229,474	T	501	563	0.69	0.66	1.13 (0.92–1.39)	0.252	1.06 (5)	0.67
10	*MINPP1*	rs12771728	89,345,292	C	850	541	0.19	0.19	1.03 (0.83–1.27)	0.788	1.06 (3)	0.34
10	*CTBP2*	rs4962416	126,686,861	G	1,202	1,388	0.17	0.17	0.99 (0.85–1.16)	0.933	1.18 (3)	0.27
11	*11q13*	rs12418451	68,691,994	A	862	876	0.12	0.13	0.94 (0.76–1.16)	0.539	1.16 (6)	0.28
11	*11q13*	rs7931342	68,751,072	G	2,445	2,018	0.79	0.77	1.15 (1.03–1.29)	0.014	1.19 (2)	0.49
11	*11q13.2*	rs10896449	68,751,242	G	2,056	1,898	0.70	0.68	1.12 (1.01–1.24)	0.031	1.17 (3)	0.28
12	*Chr. 12*	rs902774	51,560,170	A	849	541	0.09	0.09	1.02 (0.77–1.34)	0.914	1.03 (2)	0.15
15	*IL16*	rs4072111	79,365,193	T	857	539	0.04	0.05	0.87 (0.58–1.31)	0.509	1.29 (3)	0.11
16	*CDH13*	rs4782726	81,258,833	A	849	535	0.32	0.32	0.99 (0.83–1.19)	0.929	1.20 (3)	0.18
17	*HNF1B*	rs11649743	33,149,091	C	862	874	0.94	0.93	1.10 (0.83–1.45)	0.505	1.28 (7)	0.8
17	*HNF1B*	rs4430796	33,172,152	T	3,112	2,911	0.35	0.33	1.08 (1.00–1.18)	0.053	1.22 (8)	0.49
17	*HNF1B*	rs7501939	33,175,268	C	1,305	1,513	0.50	0.50	0.97 (0.87–1.09)	0.651	1.19 (8)	0.58
19	*KLK2/3*	rs887391	46,677,463	T	851	875	0.50	0.49	1.03 (0.90–1.18)	0.671	1.13 (9)	0.76
19	*KLK2/3*	rs2735839	56,056,434	A	2,773	2,680	0.32	0.31	1.07 (0.99–1.18)	0.093	1.20 (2)	0.15
22	*22q13*	rs9623117	38,782,064	C	987	1,077	0.76	0.76	0.99 (0.86–1.15)	0.937	1.25 (40)	0.21
X	*NUDT10/11*	rs5945572	51,246,422	A	1,764	1,235	0.31	0.28	1.11 (1.02–1.20)	0.02	1.23 (1)	0.35
X	*NUDT10/11*	rs5945619	51,258,411	G	1,390	1,845	0.41	0.36	1.09 (1.00–1.18)	0.039	1.19 (2)	0.36

Abbreviation: MAF, minor allele frequency.

[a]aOR: per allele OR adjusting for age and study centers.

[b]OR (per allele OR) in EDM were estimated from: (a) estimations in the original publications whenever they are available (1,2,5,6,7,8,9,19), (b) recalculations based on number of genotype counts in cases and controls when available(4), (c) Specifically for SNPs reported in Thomas et al. (3), ORs for heterozygous carriers were used as a proxy for aORs when OR for homozygous carriers were higher than OR for heterozygous carriers. For SNPs, rs4072111 and rs4782726, increases in prostate cancer risk were only seen in homozygous carriers. Therefore, the square root of OR in homozygous carriers were used as estimates.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

not only known geographic differences in ADM admixture (46), but also the different ancestry marker panels and methods used to estimate the ancestry proportions across centers (Supplementary Table 5). Therefore, we have performed all analyses with adjustment for center effects to reduce the impact of different ancestry marker panels and methods used across centers. Among those centers with ancestry marker data, inclusion of percent non-African ancestry did not substantially change the associations or inferences for any locus compared with models adjusted only for age and center.

We also evaluated the effect of the GWAS SNPs studied here on prostate cancer aggressiveness by repeating the analysis with stratification by clinical (TNM) stage and histologic (Gleason) grade (Supplementary Table 4). For SNPs that showed a significant association in the comparisons of both high grade/stage against controls and low grade/stage against controls, there were no statistically significant differences between high- and low-grade/stage cases. A number of loci were associated with disease aggressiveness, but in no instance was there an evidence for statistically significant differences in the associations by disease aggressiveness after correction for multiple testing (Supplementary Table 4).

We also evaluated whether there was evidence for first-order interactions between any of the loci identified as having a statistically significant main effect on risk of prostate cancer (Table 1). Using an additive (per-allele) model adjusted for age and study center, we considered interactions only among SNPs not in LD. The most significant interaction identified was between 2 SNPs on chromosome 8q24: rs10086908 (centromeric to Region 2) and rs6983267 (Region 3; nominal P value = 0.021). However, after correction for multiple testing using the false discovery rate (FDR), this interaction was no longer significant (FDR P value = 0.42). No other P values for interaction reached statistical significance.

Finally, we evaluated whether there was evidence for heterogeneity in associations across centers by generating forest plots of the individual center OR estimates that reached overall statistical significance (Supplementary Fig. 1). With very few exceptions, the associations that reached any level of significance showed remarkable consistency in the direction of the risk estimates. There was no statistically significant heterogeneity in effects across centers ($P > 0.05$ for all SNPs).

Discussion

A number of recent reports have modeled the role of genomic markers on prostate cancer susceptibility (1–9). We have validated a number of these loci, including 8q24, *JAZF1*, *MSMB*, 11q13, and *NUDT10/11*. In general, the point estimates of risk at these loci in our current pooled analysis of 19 studies suggest that the effects of these loci in ADM are similar to those in EDM. We also observed no statistically significant heterogeneity of effects across studies (Supplementary Fig. 1). A number of loci were not validated in our analysis, despite reaching genome-wide significance in GWAS of EDM. This discrepancy may be explained in a number of ways. First, the present study may not have been powered to identify very small effects of these loci. However, for a number of loci, we estimated Ors < 1.0 with 95% CIs that do no overlap the OR estimates originally reported in EDM. The effects of most remaining nonsignificant associations were obtained with OR < 1.05, which are lower than those estimated in EDM. If the effects of these alleles are in fact smaller in magnitude in ADM than those reported in EDM, the present study may not have been able to detect these effects. Second, allele frequencies in EDM and ADM differ at many of the loci studied here (Table 1), as do patterns of linkage disequilibrium by ethnicity (47). These differences also may affect the ability to detect significant effects at some loci in ADM, where they may have been detectable in EDM. However, the reverse situation is also possible (Table 1). Finally, if none of these limitations applies, it is possible that the loci not validated in the present study confer susceptibility only in EDM, but not ADM. Although it is unlikely that there are substantial biological differences in prostate cancer etiology between EDM and ADM, interactions of environmental exposures, prostate cancer screening, and other nongenetic risk factors may influence the penetrance of these alleles that may manifest in different risk profiles.

One of the more consistent associations identified to date is that of rs10993994 at *MSMB* (10q11; refs. 2, 3), which is confirmed as a prostate cancer susceptibility locus in ADM in this study. *MSMB* is a microseminoprotein beta gene that encodes PSP94, a nonglycosylated, cysteine-rich protein that is a member of the immunoglobulin-binding factor family synthesized by epithelial cells in the prostate and secreted into seminal plasma (3). Although the exact function of PSP94 is not well established, it is postulated to be involved in growth regulation, gene expression, and apoptosis in prostate cancer cells (2). PSP94 and its binding protein in serum, PSPBP, are potential serum markers for both prostate cancer risk and aggressiveness (48, 49), unlike the current prostate-specific antigen (PSA) screening which mainly detects the presence of prostate cancer (48). The effect of rs10993994 in *MSMB* gene expression has been investigated in function studies (5, 40). The prostate cancer risk–associated T allele of the rs10993994 SNP had only 13% of the promoter activity compared with the C allele, and treatment with increasing concentrations of the synthetic androgen R1881 resulted in a dose-dependent increase in promoter activity of the C, but not the T allele of the this SNP. In addition, tumor cell lines with a CC or CT genotype revealed a high level of *MSMB* gene expression compared with cell lines with a TT genotype. These findings were specific to the alleles of rs10993994 and not from other SNPs in the proximal promoter of *MSMB*. The significant association found in rs10993994 and lack of association found in 2 other *MSMB* SNPs included in our study also suggests the potential of rs10993994 as the causal SNP.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

Further fine-mapping studies that take advantage of the shorter LD pattern in ADM would serve to augment this hypothesis.

JAZF1 ("juxtaposed with another zinc finger protein 1") was identified by the Cancer Genetic Markers of Susceptibility (CGEMS) study as associated with prostate cancer case–control status (3). This same group has undertaken fine mapping at this locus and confirmed that the original GWAS association with rs10486567 (the SNP validated in ADM here) is likely to be the marker responsible for the association signal at this locus (50). Because rs10486567 lies in intron 2 of *JAZF1* and is not known to alter any apparent splicing or expression of this gene, the functional significance of this association has yet to be determined. JAZF1 has been associated with somatic fusion proteins in endometrial tumors (51–54), but no other genomic associations have been reported.

Two previous studies (19, 21) suggested that *NUDT10/11* was associated with prostate cancer in ADM. One study of ADM, not included in the present data, also reported that SNPs at 11q13 were associated with prostate cancer in ADM (21). The marginal association between these 2 loci and prostate cancer in this study is suggestive of validation with GWAS associations in European descent populations, but additional data may be required to fully validate these associations in ADM.

We have also validated the previously reported associations of multiple regions of chromosome 8q24 and prostate cancer in ADM. Originally identified by admixture mapping methods and GWAS (18), this locus has been shown to be composed of a number of independent prostate cancer susceptibility regions (11, 42, 55, 56). Multiple regions have been validated in our study, with the strongest association signals seen in regions 2 and 3, and our findings are consistent with the fine mapping of the admixture scan (11). The association signals seen in regions 1, 4, and a region centromeric to region 2 are much weaker compared with those in regions 2 and 3.

Finally, a number of other loci did not reach statistical significance in any analysis, and in fact provided no evidence for association with prostate cancer in ADM. These included many loci that reached genome-wide levels of significance in EA but had P value > 0.2 (and many with $P > 0.9$) in ADM (Table 1). These include associations that were reported by 2 studies of ADM that are included in the present analysis, but did not reach statistical significance in the current combined data set, including *KLK2/3* and *HNF1B/TCF2* (19, 20).

It is possible that a number of these statistically nonsignificant associations were underpowered in the present sample, especially those based on loci with lower minor allele frequencies. However, the adjusted OR estimates in ADM were often substantially lower than those reported in EA men (Table 1). Indeed, some risk estimates in ADM that had been estimated to be OR > 1 were estimated in ADM to be OR < 1, suggesting no evidence for a comparable association in between the 2 groups. There are a number of possible explanations for these findings. First, the loci identified in GWAS studies of EDM populations could represent false-positive associations that cannot be replicated in ADM. Given the large sample sizes in replication studies and strong P values associated with these loci in previous reports, this is an unlikely scenario. Second, there may be real heterogeneity in prostate cancer etiology that may be reflected by differences in allele frequency (i.e., ability to detect associations) or differences in the context in which these alleles are acting in EDM versus ADM due to differences in environmental exposures, lifestyle, or other effect modifiers not measured in studies to date. The present data do not allow us to address whether prostate cancer in ADM is less strongly influenced by genes relative to other factors than in EDM. However, the present results should be considered in future studies that may attempt to address this hypothesis. Third, the causal variants may not have been identified and genotyped yet, and the causal variants may be different in EDM and ADM. This question cannot be resolved by the data presented here and will require additional fine-mapping studies as well as ADM-specific GWAS studies in which existing GWAS loci may be validated and new loci may be identified.

Despite the validation of some prostate cancer loci in ADM, there was no strong evidence that these loci had different effects on advanced (e.g., high stage or grade) disease compared with less advanced disease (e.g., low stage or grade). This may, in part, be due to the limited power to detect significant differences between men with more versus less aggressive disease features. In some cases, there were suggestions that some SNPs were associated with more aggressive disease, including a number of SNPs at Chr. 8q24 (rs6981122, rs7000448, rs16901896) as well as others such as rs7904463 (Chr. 10) and rs5945572 (Chr. X). In these cases, there is a suggestion of stronger associations in more versus less aggressive disease in a case–control study design, but there were no statistically significant differences observed between more and less aggressive cases in a case–case comparison. Similarly, there were a number of loci for which the association was stronger for less advanced disease compared with more advanced disease. These included the associations for rs9623117 at 22q13, *MSMB* and *JAZF1* SNPs, for which the overall significant association among all cases combined (Table 1) appeared to exist only in cases with less aggressive features (Supplementary Table 4). Our results in ADM are consistent with the report by Kader et al. (57) that showed the majority of currently identified GWAS risk-associated SNPs could not differentiate aggressive from less aggressive diseases in EDM. However, contrary to the significant finding in this report showing that SNPs in *KLK2/3* and *MSMB*, both related to serum PSA levels, were associated with less aggressive disease; our null finding in *KLK2/3* and *MSMB* implies that PSA screening may not introduce the same degree of bias in cancer detection in ADM as seen in EDM.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

In studying an admixed population of ADM men, there is a concern for potential bias due to confounding by ethnicity (i.e., population stratification). To address the potential that there is bias in the risk estimates, we undertook a subset analysis of those centers that had genotyped ancestry markers and estimated the proportion of African ancestry. We observed no substantial bias in the estimates of association for any SNP. In fact, compared with associations adjusted only for age and center, the odds ratios for 7 of 47 (15%) of associations adjusted for age, center, and percent non-African ancestry changed by 5% or more: 3 of these estimates moved away from the null hypothesis whereas 4 of these estimates changed toward the null. These empirical data suggest that the potential for bias due to population stratification is not large, and that the direction of this bias may not always be away from the null hypothesis. None of these SNPs was significantly associated with the probability of having prostate cancer before or after adjustment for ancestry, so the consideration of ancestry did not change any inferences based on our results. Limitations of the approach used here include the use of different sets of markers and approaches to estimating African ancestry in only a subset of the available studies. However, our data provide no evidence for substantial bias due to population stratification in associations of GWAS SNPs in prostate cancer etiology.

In conclusion, we have validated in ADM, the associations of some, but not all, prostate cancer susceptibility loci originally identified in non-African descent populations. The finding that the genetic etiology of prostate cancer may be different in ADM and EDM suggests that studies that take advantage of the shorter LD blocks in ADM or more complete resequencing efforts will facilitate identification of causal variants in verified risk loci.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

Prostate Cancer Genetics Studies (CAP Genes) at the University of California, San Francisco: National Institutes of Health Grants CA88164, CA127298.

Flint Men's Health study (FMHS): This study was funded by a National Institutes of Health Specialized Project of Research Excellence grant in Prostate Cancer; grant number: P50CA69568. We thank Anna Ray, Dr. Ethan Lange, Kimberly Zuhlke, Joe Washburn, and the University of Michigan CDNA/Microarray Core for their help with this project.

Fox Chase Cancer Center (FCCC) Prostate Risk Assessment Program (PRAP): PA Department of Health Grant 98-PADMOH-ME-98155. We are grateful to all participants of the Prostate Cancer Risk Assessment Program at Fox Chase Cancer Center.

Fred Hutchinson Cancer Research Center (FHCRC) Prostate Cancer Studies: NIH grants R01 CA056678 and R01 CA092579 to J.L. Stanford, with additional support from the Fred Hutchinson Cancer Research Center and the National Human Genome Research Institute.

Gene-Environment Interaction in Prostate Cancer (GECAP): This study was supported by NIH grant R01- ES011126. The authors thank the GECAP study staff for their help in recruiting cases and controls, data processing and management: K. Amend, M. Aubuchon, M. Beavers, K. Bohn, J. Broderick, J. Clayton, A. Jolly, J. Mitchell, R. Rose, and D. Thomas. The authors also thank the Medical Genetics Laboratory staff for DNA processing: N. Ballard, M. McDaniel.

MD Anderson Cancer Center (MDACC): RO1CA68578, DAMD W81XWH-07-1-0645-01, P50-CA140388

Moffitt Cancer Center: Thomas Sellers, Julio Pow-Sang, Hyun Y. Park, Selina Radein, Maria Rincon, Babu Zachariah. The Moffitt group was supported by the National Cancer Institute (R01CA128813, PI: J.Y. Park).

Multiethnic Cohort (MEC) Study: Supported by National Cancer Institute (NCI) grants CA63464 and CA54281.

NCI Prostate Tissue Study (NCIPTS): Supported by the NCI intramural funding.

San Francisco Area Prostate Cancer Study (SFAPCS): Grant 99-00527V-10182 from the California Cancer Research Fund.

Los Angeles County Study (LACS): Funded by grant 99-00524V-10258 (to SAI) from the Cancer Research Fund, under Interagency Agreement 97-12013 (University of California contract 98-00924V) with the Department of Health Services Cancer Research Program and by grant R01CA84979 (to SAI) from the National Cancer Institute, National Institutes of Health.

UK Genetic Prostate Cancer Study (UKGPCS): This work was supported by Cancer Research UK Grant C5047/A7357. We thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. We acknowledge the NCRN nurses and Consultants for their work in the UKGPCS study.

Prostate Cancer Clinical Outcome Study (PC2OS) at the University of Louisville: The authors thank Tiva Van Cleave and Nicole Lavender for data collection and analysis. Rick A. Kittles for generously donating DNA samples collected from men of African descent. This study was supported in part by the NIH R03 CA128028 and the James Graham Brown Cancer Center and the Bucks for Brains "Our Highest Potential" in Cancer Research Endowment.

University of Pennsylvania Study of Clinical Outcomes, Risk, and Ethnicity (SCORE): This work was supported by R01-CA08574 and P01-CA105641. The authors thank Drs. D. Goldmann, W. Greer, G.W. Crooks, D.A. Horowitz, D. Farhadi, M.D. Cirigliano, M. Rusk, V. Weil, S.J. Gluckman, C. Bridges, M.L. Walker, and C. Guerra for their invaluable assistance in ascertaining study participants.

University of Texas San Antonio Center for Biomarkers of Risk for Prostate Cancer (SABOR): SABOR is a Clinical and Epidemiologic Center of the Early Detection Research Network of the National Cancer Institute, supported by NIH UO1 CA86402. Prevalence samples obtained through support from the American Cancer Society (TURSG-03-152-01-CCE). The authors thank Dr. Joke Beuten for assistance with these genetic studies.

Wake Forest Consortium: The work at Washington University was supported by NCI R01CA112028, the St. Louis Men's Group against Cancer and the Par for the Cure Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 1, 2010; revised September 14, 2010; accepted September 21, 2010; published OnlineFirst November 11, 2010.

References

1. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, et al. (2008). Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet 40, 281-283.
2. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40, 316-321.
3. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40, 310-315.
4. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, et al. (2007). Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst 99, 1836-1844.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

5. Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, et al. (2009). Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc Natl Acad Sci U S A 106, 7933-7938.
6. Zheng SL, Stevens VL, Wiklund F, Isaacs SD, Sun J, Smith S, et al. (2009). Two independent prostate cancer risk-associated Loci at 11q13. Cancer Epidemiol Biomarkers Prev 18, 1815-1820.
7. Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, Gao Z, et al. (2008). Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet 40, 1153-1155.
8. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. (2007). Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 39, 977-983.
9. Hsu FC, Sun J, Wiklund F, Isaacs SD, Wiley KE, Purcell LD, Gao, et al. (2009). A Novel Prostate Cancer Susceptibility Locus at 19q13. Cancer Res 69, 2720–2723.
10. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, et al. (2009). Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet 41, 1058-1060.
11. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, et al. (2007). Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 39, 638-644.
12. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, et al. (2009). Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet 41, 1116-1121.
13. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, et al. (2009). Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet 41, 1122-1126.
14. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39, 645-649.
15. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. (2006). A common variant associated with prostate cancer in European and African populations. Nat Genet 38, 652-658.
16. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, et al. (2009). Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet 41, 1055-1057.
17. Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, Ingles SA, et al. (2008). Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. Cancer Epidemiol Biomarkers Prev 17, 2052-2061.
18. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, et al. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 103, 14068-14073.
19. Waters KM, Le Marchand L, Kolonel LN, Monroe KR, Stram DO, Henderson BE, et al. (2009). Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. Cancer Epidemiol Biomarkers Prev 18, 1285-1289.
20. Liu X, Plummer SJ, Nock NL, Casey G, and Witte JS. (2006). Nonsteroidal antiinflammatory drugs and decreased risk of advanced prostate cancer: modification by lymphotoxin alpha. Am J Epidemiol 164, 984-989.
21. Stanford JL, Wicklund KG, McKnight B, Daling JR, and Brawer MK, et al. (1999). Vasectomy and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 8, 881-886.
22. Agalliu I, Salinas CA, Hansten PD, Ostrander EA, and Stanford JL, et al. (2008). Statin use and risk of prostate cancer: results from a population-based epidemiologic study. Am J Epidemiol 168, 250-260.
23. Giri VN, Beebe-Dimmer J, Buyyounouski M, Konski A, Feigenberg SJ, Uzzo RG, et al. (2007). Prostate cancer risk assessment program: a 10-year update of cancer detection. J Urol 178, 1920-1924; discussion 1924.
24. Cooney KA, Strawderman MS, Wojno KJ, Doerr KM, Taylor A, Alcser KH, et al. (2001). Age-specific distribution of serum prostate-specific antigen in a community-based study of African-American men. Urology 57, 91-96.
25. Heeringa SG, Alcser KH, Doerr K, Strawderman M, Cooney K, Medbery B, and Schottenfeld D. (2001). Potential selection bias in a community-based study of PSA levels in African-American men. J Clin Epidemiol 54, 142-148.
26. Rybicki BA, Neslund-Dudas C, Nock NL, Schultz LR, Eklund L, Rosbolt J, et al. (2006). Prostate cancer risk from occupational exposure to polycyclic aromatic hydrocarbons interacting with the GSTP1 Ile105Val polymorphism. Cancer Detect Prev 30, 412-422.
27. Schwartz G, John E, Rowland G, and Ingles S. (2010). Schwartz GG, John EM, Rowland G Ingles SA. Prostate cancer in African American men and polymorphism in the calcium sensing receptor. Cancer Biology & Therapy (in press).
28. Benford ML, VanCleave TT, Lavender NA, Kittles RA, and Kidd LR. (2010) 8q24 sequence variants in relation to prostate cancer risk among men of African descent: a case-control study. BMC Cancer 10, 334.
29. Strom SS, Yamamura Y, Flores-Sandoval FN, Pettaway CA, and Lopez DS. (2008). Prostate cancer in Mexican-Americans: identification of risk factors. Prostate 68, 563-570.
30. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. (2000). A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 151, 346-357.
31. Park JY, Tanner JP, Sellers TA, Huang Y, Stevens CK, Dossett N, et al. (2007). Association between polymorphisms in HSD3B1 and UGT2B17 and prostate cancer risk. Urology 70, 374-379.
32. Rebbeck TR, Jaffe JM, Walker AH, Wein AJ, and Malkowicz SB. (1998). Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. [see comment] [erratum appears in J Natl Cancer Inst 1999 Jun 16;91(12):1082]. Journal of the National Cancer Institute 90, 1225-1229.
33. Beuten J, Gelfond JA, Franke JL, Weldon KS, Crandall AC, Johnson-Pais TL, et al. (2009). Single and multigenic analysis of the association between variants in 12 steroid hormone metabolism genes and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 18, 1869-1880.
34. Beuten J, Gelfond JA, Martinez-Fierro ML, Weldon KS, Crandall AC, Rojas-Martinez A, et al. (2009). Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men. Carcinogenesis 30, 1372-1379.
35. John EM, Schwartz GG, Koo J, Van Den Berg D, and Ingles SA. (2005). Sun exposure, vitamin D receptor gene polymorphisms, and risk of advanced prostate cancer. Cancer Res 65, 5470-5479.
36. Xu J, Kibel AS, Hu JJ, Turner AR, Pruett K, Zheng SL, et al. (2009). Prostate cancer risk associated loci in African Americans. Cancer Epidemiol Biomarkers Prev 18, 2145-2149.
37. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 39, 631-637.
38. Hsu FC, Sun J, Wiklund F, Isaacs SD, Wiley KE, Purcell LD, et al. (2009). A novel prostate cancer susceptibility locus at 19q13. Cancer Res 69, 2720-2723.
39. Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, et al. (2009). Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. Hum Mol Genet 18, 1368-1375.
40. Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, et al. (2009). Sequence variants at 22q13 are associated with prostate cancer risk. Cancer Res 69, 10-15.
41. Zheng SL, Sun J, Cheng Y, Li G, Hsu FC, Zhu Y, et al. (2007). Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. J Natl Cancer Inst 99, 1525-1533.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-575.
43. Haploview. http://www.broad.mit.edu/mpg/haploview
44. Salinas CA, Kwon E, Carlson CS, Koopmeiners JS, Feng Z, Karyadi DM, Ostrander EA, and Stanford JL. (2008). Multiple independent genetic variants in the 8q24 region are associated with prostate cancer risk. Cancer Epidemiol Biomarkers Prev 17, 1203-1213.

Published OnlineFirst November 11, 2010; DOI:10.1158/1055-9965.EPI-10-0698

45. Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, et al. (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol 114, 18-29.
46. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. (2001). Linkage disequilibrium in the human genome. Nature 411, 199-204.
47. Nam RK, Reeves JR, Toi A, Dulude H, Trachtenberg J, Emami M, et al. (2006). A novel serum marker, total prostate secretory protein of 94 amino acids, improves prostate cancer detection and helps identify high grade cancers at diagnosis. J Urol 175, 1291-1297.
48. Reeves JR, Dulude H, Panchal C, Daigneault L, and Ramnani DM. (2006). Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. Clin Cancer Res 12, 6018-6022.
49. Yi-Ping F, Wei T, Jacobs K, Hayes R, Kraft P, Berndt S, et al. (2010). Refining the prostate cancer genetic association within the JAZF1 gene on chromosome 7p15.2. Cancer Epidemiol Biomarkers Prev 19, 1349–1355.
50. Koontz JI, Soreng AL, Nucci M, Kuo FC, Pauwels P, van Den Berghe H, et al. (2001). Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors. Proc Natl Acad Sci U S A 98, 6348-6353.
51. Micci F, Walter CU, Teixeira MR, Panagopoulos I, Bjerkehagen B, Saeter G, and Heim S. (2003). Cytogenetic and molecular genetic analyses of endometrial stromal sarcoma: nonrandom involvement of chromosome arms 6p and 7p and confirmation of JAZF1/JJAZ1 gene fusion in t(7;17). Cancer Genet Cytogenet 144, 119-124.
52. Micci F, Panagopoulos I, Bjerkehagen B, and Heim S. (2006). Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. Cancer Res 66, 107-112.
53. Oliva E, de Leval L, Soslow RA, and Herens C. (2007). High frequency of JAZF1-JJAZ1 gene fusion in endometrial stromal tumors with smooth muscle differentiation by interphase FISH detection. Am J Surg Pathol 31, 1277-1284.
54. Hooker S, Hernandez W, Chen H, Robbins C, Torres JB, Ahaghotu C, et al. Replication of prostate cancer risk loci on 8q24, 11q13, 17q12, 19q33, and Xp11 in African Americans. Prostate 70, 270-275.
55. Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, Kolonel LN, et al. (2007). A common genetic risk factor for colorectal and prostate cancer. Nat Genet 39, 954-956.
56. Robbins C, Torres JB, Hooker S, Bonilla C, Hernandez W, Candreva A, et al. (2007). Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. Genome Res 17, 1717-1722.
57. Kader AK, Sun J, Isaacs SD, Wiley KE, Yan G, Kim ST, et al. (2009). Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. Prostate 69, 1195-1205.

nature genetics

Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21

Christopher A Haiman[1], Gary K Chen[1], William J Blot[2,3], Sara S Strom[4], Sonja I Berndt[5], Rick A Kittles[6], Benjamin A Rybicki[7], William B Isaacs[8], Sue A Ingles[1], Janet L Stanford[9], W Ryan Diver[10], John S Witte[11], Ann W Hsing[5], Barbara Nemesure[12], Timothy R Rebbeck[13], Kathleen A Cooney[14], Jianfeng Xu[15], Adam S Kibel[16], Jennifer J Hu[17], Esther M John[18], Serigne M Gueye[19], Stephen Watya[20], Lisa B Signorello[2,3], Richard B Hayes[21], Zhaoming Wang[5], Edward Yeboah[22], Yao Tettey[22], Qiuyin Cai[3], Suzanne Kolb[9], Elaine A Ostrander[23], Charnita Zeigler-Johnson[13], Yuko Yamamura[4], Christine Neslund-Dudas[7], Jennifer Haslag-Minoff[16], William Wu[16], Venetta Thomas[17], Glenn O Allen[17], Adam Murphy[24], Bao-Li Chang[13], S Lilly Zheng[15], M Cristina Leske[12], Suh-Yuh Wu[12], Anna M Ray[14], Anselm J M Hennis[12,25,26], Michael J Thun[10], John Carpten[27], Graham Casey[1], Erin N Carter[1], Edder R Duarte[1], Lucy Y Xia[1], Xin Sheng[1], Peggy Wan[1], Loreall C Pooler[1], Iona Cheng[28], Kristine R Monroe[1], Fredrick Schumacher[1], Loic Le Marchand[28], Laurence N Kolonel[28], Stephen J Chanock[5], David Van Den Berg[1], Daniel O Stram[1] & Brian E Henderson[1]

In search of common risk alleles for prostate cancer that could contribute to high rates of the disease in men of African ancestry, we conducted a genome-wide association study, with 1,047,986 SNP markers examined in 3,425 African-Americans with prostate cancer (cases) and 3,290 African-American male controls. We followed up the most significant 17 new associations from stage 1 in 1,844 cases and 3,269 controls of African ancestry. We identified a new risk variant on chromosome 17q21 (rs7210100, odds ratio per allele = 1.51, $P = 3.4 \times 10^{-13}$). The frequency of the risk allele is ~5% in men of African descent, whereas it is rare in other populations (<1%). Further studies are needed to investigate the biological contribution of this allele to prostate cancer risk. These findings emphasize the importance of conducting genome-wide association studies in diverse populations.

Genome-wide association studies (GWAS) of prostate cancer have identified more than 30 variants associated with risk that, in aggregate, are estimated to account for approximately 20% of the familial risk of prostate cancer[1–12]. Aside from admixture and fine-mapping studies that identified multiple independent risk variants at 8q24 (refs. 13,14), and a more recent GWAS among Japanese men that identified five new loci[9], discoveries in prostate cancer have come from studies in men of European ancestry. However, prostate cancer incidence in men of African ancestry is greater than in non-African populations[15], with the disparity presumably reflecting both differences in prevalence of environmental risk factors and susceptibility alleles that are shared among men of African descent. For example, the risk variants at 8q24, many of which are more common in men of African ancestry[14], could contribute partly to the greater incidence of prostate cancer in this population and provide some support for the hypothesis of a genetic contribution underlying racial and ethnic disparities in disease risk.

We assembled a consortium of prostate cancer studies that included men of African ancestry and conducted a GWAS to search for additional risk loci that may be more common in men of African descent. Stage 1 included 3,621 African-American cases with prostate cancer and 3,502 African-American controls drawn from 11 studies (**Supplementary Table 1** and Online Methods). We conducted genotyping in stage 1 using the Illumina Infinium 1M-Duo. Following

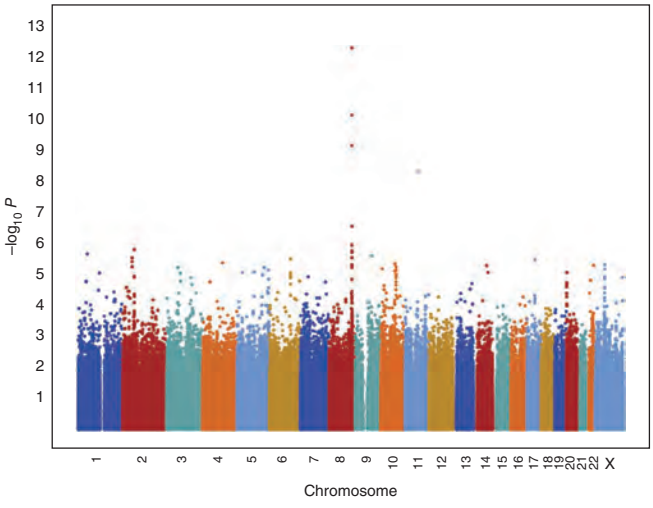


Figure 1 A plot of the $-\log_{10} P$ values by chromosome.

A full list of author affiliations appears at the end of the paper.

Received 23 February; accepted 21 April; published online 22 May 2011; doi:10.1038/ng.839

Table 1 The association of variant rs7210100 at 17q21 with prostate cancer risk in men of African ancestry

Stage 1 studies	Cases/controls[a]	RAF in controls	OR[b]	95% CI[b]	*P*[c]
MEC	1,060/1,055	0.04	1.58	1.21–2.08	8.8×10^{-4}
SCCS	201/412	0.05	1.40	0.85–2.31	0.19
PLCO	227/239	0.05	1.44	0.82–2.52	0.21
CPS-II	64/112	0.07	0.66	0.24–1.78	0.41
MDA	527/437	0.05	1.39	0.95–2.02	0.089
IPCG	354/157	0.05	1.54	0.84–2.82	0.17
LAAPC	288/287	0.06	0.94	0.57–1.56	0.81
CaP Genes	71/85	0.06	1.72	0.78–3.82	0.18
DCPD	263/341	0.07	1.14	0.75–1.75	0.54
KCPCS	141/75	0.05	0.95	0.42–2.16	0.90
GECAP	224/89	0.05	2.47	1.14–5.34	0.022
Combined	3,420/3,289		1.40	1.21–1.62	5.2×10^{-6}
					$P_{Het} = 0.89$[d]
Stage 2 studies					
SFPCS	86/36	0.04	1.86	0.53–6.55	0.34
FMHS	125/339	0.06	1.70	0.98–2.93	0.058
MEC-LAC	551/555	0.04	1.92	1.30–2.83	9.7×10^{-4}
NCPCS	214/249	0.06	0.92	0.51–1.66	0.79
WFPCS	58/65	0.04	1.90	0.56–6.42	0.30
WUPCS	73/153	0.04	1.96	0.76–5.03	0.16
GHS	264/964	0.07	1.37	0.94–2.01	0.11
Combined	1,371/2,361		1.55	1.26–1.89	2.5×10^{-5}
					$P_{Het} = 0.25$[d]
Stage 3 studies					
SCORE	146/267	0.05	1.58	0.88–2.83	0.13
PROGRÈS	79/395	0.05	2.64	1.36–5.10	4.0×10^{-3}
PCBP	246/242	0.05	2.02	1.20–3.39	7.9×10^{-3}
Combined	471/904		2.07	1.49–2.88	1.5×10^{-5}
					$P_{Het} = 0.51$[d]
Stages 1+2+3	**5,262/6,554**		**1.51**	**1.35–1.69**	**3.4×10^{-13}**
					$P_{Het} = 0.58$[d]

[a]Number of cases and controls with genotype data for rs7210100. [b]Adjusted for age and eigenvectors 1–10 in stage 1 (and study in pooled analysis). Adjusted for age in stage 2 and stage 3. Adjusted for age and study in stage 1+2+3 analysis. [c]P for trend (1 degree of freedom). [d]Test of heterogeneity. RAF, risk allele frequency; OR, odds ratio; 95% CI, 95% confidence interval.

quality-control exclusions (Online Methods), the stage 1 analysis consisted of 1,047,986 SNPs (minor allele frequency ≥0.01) examined in 3,425 cases and 3,290 controls.

In comparing (for all SNPs) the observed with the expected distribution of *P* values from a 1-degree-of-freedom trend test, there was evidence of inflation in the test statistic ($\lambda = 1.11$). Principal components analysis highlighted the high degree of admixture in this population, and the overinflation diminished following additional adjustment for ancestry ($\lambda = 1.03$; **Supplementary Fig. 1** and Online Methods). The association of four SNPs achieved genome-wide significance in the stage 1 sample, with *P* values between $P = 5.4 \times 10^{-9}$ and $P = 5.7 \times 10^{-13}$ (**Fig. 1**). These SNPs are located in known prostate cancer risk regions, three of which are at 8q24 (rs10505483, rs1456315 and rs7824364 at 128.173–128.205 Mb (NCBI36) and one of which is at 11q13 (rs7130881 at 67.75 Mb).

We selected 17 SNPs ($P < 2 \times 10^{-5}$) located outside of known prostate cancer risk regions to examine in a second stage. The associations of these 17 SNPs with prostate cancer risk were not influenced substantially by population stratification in the stage 1 sample as evaluated by principal components analysis (**Supplementary Table 2**). The stage 2 sample included 1,396 cases and 2,383 controls of African ancestry from seven independent studies: six US-based studies and one study in Ghana. Of the 17 SNPs, only marker rs7210100 at 17q21 was significantly associated with risk in the stage 2 studies (odds ratio (OR) = 1.55, $P = 2.5 \times 10^{-5}$; **Table 1**). None of the other SNPs selected in stage 1 were significantly associated with risk in the stage 2 sample (all *P* values were >0.05); we excluded rs13116912 because it deviated from Hardy-Weinberg equilibrium in the majority of stage 2 studies. The results for all 17 SNPs in stage 1 and stage 2 are presented in **Supplementary Table 3**.

We further examined the association with rs7210100 in a third stage that included three studies among men of African descent, a study from the United States (SCORE), a study in Senegal (PROGRÈS) and a study in Barbados (PCBP). We found rs7210100 to be positively associated with risk in all three studies (stage 3, 471 cases and 904 controls; combined OR = 2.07, $P = 1.5 \times 10^{-5}$; **Table 1**).

Adjustment for global ancestry or local ancestry (African versus European) in the stage 1 studies did not influence the results for rs7210100 (OR = 1.41 without adjustment for ancestry, OR = 1.40 adjusted for global ancestry and OR = 1.43 adjusted for global and local ancestry). The effect estimate for rs7210100 was also similar in men with <15% global European ancestry (1,251 cases and 1,325 controls; OR = 1.41) as well as in cases and controls estimated to have two chromosomes of African ancestry at this location (2,214 cases and 2,080 controls; OR = 1.47). We observed no evidence of heterogeneity of the association by study for this variant in the stage 1 ($P_{het} = 0.89$), stage 2 ($P_{het} = 0.25$) or stage 3 studies ($P_{het} = 0.51$) or among all studies ($P_{het} = 0.58$). Results for all SNPs examined in the replication stages were also unaffected when adjusting for European ancestry in studies in which information on global ancestry was available (**Supplementary Tables 4** and **5**).

In combining the results across all three stages (5,262 cases and 6,554 controls), rs7210100 was strongly and significantly associated with risk (OR = 1.51, 95% CI 1.35–1.69, $P = 3.4 \times 10^{-13}$). The risk for heterozygote and homozygote carriers was 1.49 (95% CI 1.32–1.68) and 2.73 (95% CI 1.50–4.96), respectively. We did not find any stronger signal with imputed SNPs to the phase 2 HapMap populations in the surrounding region at chromosome 17q21 (**Fig. 2** and **Supplementary Fig. 2**).

The association with rs7210100 was similar when stratifying on age ($P = 0.72$) and first-degree family history of prostate cancer ($P = 0.36$). We also observed no significant difference in the association of rs7210100 with prostate cancer stage ($P = 0.94$) or tumor grade ($P = 0.11$) at diagnosis. However, the association with rs7210100 was greater for non-advanced disease when classified based on stage and grade (Gleason score <8 and localized stage, 2,433 cases and 6,554 controls, OR = 1.67, $P = 8.6 \times 10^{-12}$) than for advanced disease (Gleason score ≥8 or non-localized disease, 1,719 cases and 6,554 controls, OR = 1.27, $P = 5.0 \times 10^{-3}$, $P_{het} = 6.0 \times 10^{-3}$).

Among controls with prostate-specific antigen (PSA) levels measured at ≤4 ng/ml ($n = 2,383$), we found no significant association

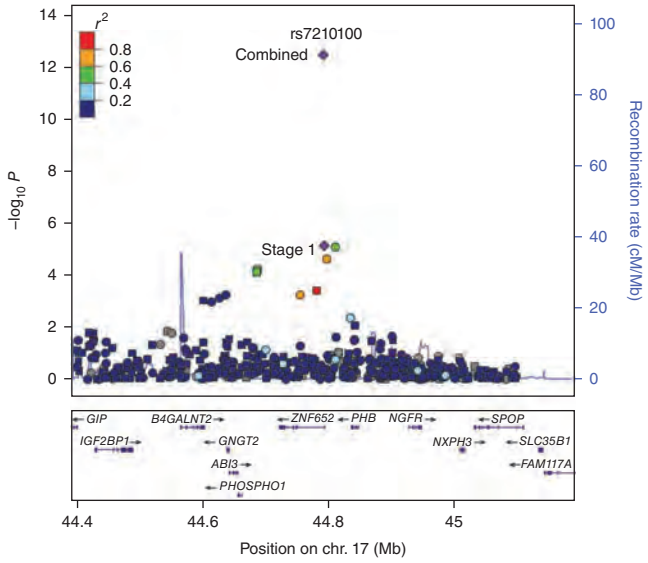


Figure 2 A regional plot of the $-\log_{10} P$ values for genotyped (squares) and imputed (circles) SNPs at the chromosome 17q21 risk locus in the stage 1 African-American sample. The shading depicts the strength of the correlation (r^2) between rs7210100 and the SNPs tested in the region. The correlation is estimated in the YRI population from the 1000 Genomes Project (June 2010). Also shown are human genome build 18 coordinates (Mb), recombination rates in cM per Mb and genes in the region. The plot was generated using LocusZoom.

between PSA levels and rs7210100 genotype (P = 0.58). Limiting the analysis to controls with PSA levels <4 ng/ml and cases from these studies did not change the association between rs7210100 and prostate cancer risk (n = 3,157 cases and 2,383 controls, OR = 1.62, $P = 4.5 \times 10^{-8}$).

The variant rs7210100 is located in intron 1 of *ZNF652* on chromosome 17q21.32. *ZNF652* encodes a zinc-finger protein transcription factor that has been shown to interact with the eight-twenty-one (ETO) protein, CBFA2T3, which acts as a transcriptional repressor by forming complexes with co-repressor proteins and HDACs[16]. Co-expression of ZNF652 and the androgen receptor in prostate tumors has been associated with a decrease in relapse-free survival[17]. A common variant just upstream of *ZNF652* has also been associated with blood pressure in a GWAS of men and women of European ancestry[18]. Sequencing of the five coding exons of *ZNF652* in 48 subjects (with an oversampling of risk allele carriers; Online Methods) did not reveal a coding variant strongly correlated with rs7210100. Further work is needed to map this locus in order to nominate optimal candidate markers, in addition to rs7210100, for functional studies in pursuit of regulatory effects of one or more variants in the region.

The risk allele of rs7210100 is relatively uncommon in men of African ancestry (4–7%), and is extremely rare (<1%) in non-African populations as reported by the 1000 Genomes Project. The frequency of the risk allele in men of west-African ancestry (Ghana and Senegal) is very similar to that observed in African Americans, as well as in men from east Africa (Uganda; n = 111, risk allele frequency = 0.04). GWAS in populations of European ancestry have not pointed to this region of 17q21 as a risk locus for prostate cancer (**Supplementary Fig. 3**). Together, these observations suggest that the underlying biologically relevant allele may be limited to populations of African descent. As reported by the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program, prostate cancer incidence in African-American men is 1.56 times higher than the incidence in non-Hispanic individuals of European descent. Because approximately 10% of African-American men carry this variant, which increases their risk 1.50-fold over non-carriers, we estimate that this locus may be responsible for as much as 9% (95% CI 6–12%) of the greater incidence of prostate cancer in African-American men (Online Methods).

In summary, we detected a marker of risk for prostate cancer that appears specific to men of African descent, who have an increased incidence and mortality of this disease. These findings provide strong support for conducting GWAS in diverse populations to identify markers of risk that may be population specific and which could contribute to racial and ethnic disparities in disease incidence. Further work is needed to characterize the 17q21 region and conduct the functional studies required to understand the role of this germ-line variation in prostate cancer susceptibility.

URLs. SEER, http://seer.cancer.gov/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The Multiethnic Cohort (MEC) and the genotyping in this study were supported by US National Institutes of Health (NIH) grants CA63464, CA54281, CA1326792, CA148085 and HG004726. Genotyping of the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and Ghana samples was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), NIH. The Los Angeles Study of Aggressive Prostate Cancer (LAAPC) was funded by grant 99-00524V-10258 from the Cancer Research Fund under Interagency Agreement #97-12013 (University of California contract #98-00924V) with the Department of Health Services Cancer Research Program. Cancer incidence data for the MEC and LAAPC studies have been collected by the Los Angeles Cancer Surveillance Program of the University of Southern California with Federal funds from the NCI, NIH, Department of Health and Human Services under Contract No. N01-PC-35139 and the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885 and grant number 1U58DP000807-3 from the Centers for Disease Control and Prevention (CDC). The King County (Washington) Prostate Cancer Study (KCPCS) was supported by NIH grants CA056678, CA082664 and CA092579 with additional support from the Fred Hutchinson Cancer Research Center and the Intramural Program of the National Human Genome Research Institute. Prostate Cancer Case-Control Studies at MD Anderson (MDA) was support by grants CA68578, ES007784, DAMD W81XWH-07-1-0645 and CA140388. The Gene-Environment Interaction in Prostate Cancer Study (GECAP) was supported by NIH grant ES011126. Prostate Cancer Genetics Study (CaP Genes) was supported by CA88164 and CA127298. Identifying Prostate Cancer Genes (IPCG) was supported by Department of Defense (DOD) grant W81XWH-07-1-0122. Case-Control Study of Prostate Cancer among African Americans in Washington, DC (DCPC) was supported by NIH grant S06GM08016 and DOD grants DAMD W81XWH-07-1-0203 and DAMD W81XWH-06-1-0066. The Southern Community Cohort Study (SCCS) is funded by NIH grant CA092447. SCCS sample preparation was conducted at the Epidemiology Biospecimen Core Lab, which is supported in part by the Vanderbilt-Ingram Cancer Center (CA68485). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; and Arkansas Department of Health, Cancer Registry. The Arkansas Central Cancer Registry is fully funded by

a grant from National Program of Cancer Registries, CDC. Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry, which participates in the National Program of Cancer Registries (NPCR) of the CDC. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry. We thank C. Edlund from the University of Southern California (USC) Genomics Core for his technical and informatics assistance. The authors thank C. Berg and P. Prorok, Division of Cancer Prevention, NCI, the screening center investigators and staff of the PLCO Cancer Screening Trial, T. Riley and staff at Information Management Services, Inc. and B. O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial. We also acknowledge the technical support of M. Gielzak and G. Yan. The authors would like to thank E. Tay and V. Okyne for their expert help in coordinating the Ghana study; A. DeMarzo and G. Netto of Johns Hopkins University for pathology review; V. Devairakkam, N. Kim and J. Heinrich of Research Triangle Institute (RTI) for their expert study management; and S. Niwa and A. Truelove of Westat for study support and data management. The Washington University Prostate Cancer Study (WUPCS) was supported by CA112028, the St. Louis Men's Group Against Cancer and the David H. Nickerson Foundation. The San Francisco Bay Area Prostate Cancer Study (SFPCS) was funded by grant 99-00527V-10182 from the California Cancer Research Program. The Study of Clinical Outcomes, Risk and Ethnicity (SCORE) was supported by CA085074 and CA105641, as well as a grant from the Commonwealth of Pennsylvania. The Cancer Prevention Study II Nutrition Cohort (CPS-II) is supported by the American Cancer Society. Prostate Cancer in a Black Population (PCBP) was supported by grant CA114379 (NCI) and contract N01-HG-25487 (National Health Genome Research Institute and the Office of Minority Health).

AUTHOR CONTRIBUTIONS

C.A.H., D.O.S. and B.E.H. contributed to the study concept and design. L.C.P., E.N.C., E.R.D., L.Y.X., D.V.D.B., S.J.C., C.A.H. and X.S. supervised the laboratory analyses and quality control. G.K.C., P.W. and D.O.S. contributed to the statistical analysis. C.A.H. wrote the manuscript. W.J.B., S.S.S., S.I.B., R.A.K., B.A.R., W.B.I., S.A.I., J.L.S., W.R.D., J.S.W., A.W.H., B.N., T.R.R., K.A.C., J.X., A.S.K., J.J.H., E.M.J., S.M.G., S.W., L.B.S., R.B.H., Z.W., E.Y., Y.T., Q.C., S.K., E.A.O., C.Z.-J., Y.Y., C.N.-D., J.H.-M., W.W., V.T., G.O.A., A.M., B.-L.C., S.L.Z., M.C.L., S.-Y.W., A.M.R., A.J.M.H., M.J.T., J.C., G.C., I.C., K.R.M., F.S., L.L.M., L.N.K. and B.E.H. conducted the epidemiologic studies that contributed to the scan. All authors helped in the interpretation and discussion of the findings and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Al Olama, A.A. *et al.* Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat. Genet.* **41**, 1058–1060 (2009).
2. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat. Genet.* **38**, 652–658 (2006).
3. Eeles, R.A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.* **41**, 1116–1121 (2009).
4. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
5. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.* **41**, 1122–1126 (2009).
6. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
7. Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat. Genet.* **40**, 281–283 (2008).
8. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
9. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.* **42**, 751–754 (2010).
10. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
11. Yeager, M. *et al.* Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **41**, 1055–1057 (2009).
12. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
13. Freedman, M.L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. USA* **103**, 14068–14073 (2006).
14. Haiman, C.A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.* **39**, 638–644 (2007).
15. American Cancer Society. *Cancer Facts & Figures for African Americans 2009–2010.* (American Cancer Society, Atlanta, GA, 2009).
16. Amann, J.M. *et al.* ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol. Cell. Biol.* **21**, 6470–6483 (2001).
17. Callen, D.F. *et al.* Co-expression of the androgen receptor and the transcription factor ZNF652 is related to prostate cancer outcome. *Oncol. Rep.* **23**, 1045–1052 (2010).
18. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).

[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, California, USA. [2]International Epidemiology Institute, Rockville, Maryland, USA. [3]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University and the Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA. [4]Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [6]Department of Medicine, University of Illinois at Chicago, Chicago, Illinois, USA. [7]Department of Biostatistics and Research Epidemiology, Henry Ford Hospital, Detroit, Michigan, USA. [8]James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institutions, Baltimore, Maryland, USA. [9]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [10]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA. [11]Institute for Human Genetics, Departments of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA. [12]Department of Preventive Medicine, Stony Brook University, Stony Brook, New York, USA. [13]University of Pennsylvania School of Medicine and the Abramson Cancer Center, Philadelphia, Pennsylvania, USA. [14]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA. [15]Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA. [16]Division of Urology, Department of Surgery, Washington University, St. Louis, Missouri, USA. [17]Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida, USA. [18]Cancer Prevention Institute of California, Fremont, California, USA. [19]Hôpital Général de Grand Yoff, Dakar, Senegal. [20]Mulago Hospital, Makerere University, Department of Surgery, Urology Unit, Kampala, Uganda. [21]Division of Epidemiology, Department of Environmental Medicine, New York University Langone Medical Center, New York, New York, USA. [22]University of Ghana Medical School, Legon, Ghana. [23]Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. [24]Department of Urology, Northwestern University, Chicago, Illinois, USA. [25]Chronic Disease Research Centre and Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados. [26]Ministry of Health, Bridgetown, Barbados. [27]The Translational Genomics Research Institute, Phoenix, Arizona, USA. [28]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA. Correspondence should be addressed to C.A.H. (haiman@usc.edu).

ONLINE METHODS

Studies. The studies included in stage 1 were drawn from 11 epidemiological studies of prostate cancer among African-American men. These studies included The Multiethnic Cohort (MEC; 1,094 cases and 1,096 controls), The Southern Community Cohort Study (SCCS, 212 cases and 419 controls), The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO, 286 cases and 269 controls), The Cancer Prevention Study II Nutrition Cohort (CPS-II, 76 cases and 152 controls), Prostate Cancer Case-Control Studies at MD Anderson (MDA, 543 cases and 474 controls), Identifying Prostate Cancer Genes (IPCG, 368 cases and 172 controls), The Los Angeles Study of Aggressive Prostate Cancer (LAAPC, 296 cases and 303 controls), Prostate Cancer Genetics Study (CaP Genes, 75 cases and 85 controls), Case-Control Study of Prostate Cancer among African Americans in Washington, DC (DCPC, 292 cases and 359 controls), King County (Washington) Prostate Cancer Study (KCPCS, 145 cases and 81 controls) and The Gene-Environment Interaction in Prostate Cancer Study (GECAP, 234 cases and 92 controls). These studies provided DNA samples for 3,621 cases and 3,502 controls.

Stage 2 included 1,396 cases and 2,383 controls from seven studies: San Francisco Bay Area Prostate Cancer Study (SFPCS, 86 cases and 37 controls), The Flint Men's Health Study (FMHS, 135/353), The Multiethnic Cohort/Los Angeles County (MEC-LA, 554 cases and 557 controls), North Carolina Prostate Cancer Study (NCPCS, 214 cases and 249 controls), Wake Forest University Prostate Cancer Study (WFPCS, 59 cases and 66 controls), Washington University Prostate Cancer Study (WUPCS, 75 cases and 153 controls) and The Ghana Men's Health Study (GHS, 271 cases and 968 controls). Stage 3 included 484 cases and 947 controls from three studies: The Study of Clinical Outcomes, Risk and Ethnicity (SCORE, 152 cases and 280 controls), Prostate-Genetique-Recherche-Senegal (PROGRÈS, 86 cases and 414 controls) and Prostate Cancer in a Black Population (PCBP, 246 cases and 253 controls). Detailed information about the design and organization of each study is provided in the **Supplementary Note**.

Genotyping and quality control. Genotyping in stage 1 (3,621 cases and 3,502 controls) was conducted using the Illumina Infinium Human1M-Duo. Samples ($n = 408$) were removed based on the following exclusion criteria: (i) unknown replicates across studies, (ii) call rates <95%, (iii) >10% mean heterozygosity on the X chromosome and/or <10% mean intensity on the Y chromosome, (iv) ancestry outliers and (v) samples that were related (discussed below). The concordance rate for 158 replicate samples was 99.99%. Starting with 1,153,397 SNPs, we removed SNPs with <95% call rate, minor allele frequencies <1% or >1 quality-control mismatch based on sample replicates ($n = 105,411$). The analysis included 1,047,986 SNPs among 3,425 cases and 3,290 controls.

We used PLINK (see URLs) to calculate the probabilities of sharing 0, 1 and 2 alleles (Z = Z0, Z1, Z2) across all possible pairs of samples to determine individuals who were likely to be related to others within and across studies. We identified 167 pairs of related subjects (monozygotic twin, parent-offspring, full- and half-sibling pairs) based on the values of their observed probability vector Z being within 1 standard deviation of the expected values of Z for their respective relationship. The criterion for removal was such that individuals that were connected with a higher number of pairs were chosen for removal. In all other cases, one of the two members was randomly selected for removal. A total of 141 subjects were removed.

The EIGENSTRAT (see URLs) software was used to calculate eigenvectors that explained genetic differences in ancestry among samples in the study[19]. We included data from both HapMap populations (CEPH (Utah residents with ancestry from northern and western Europe) (CEU), Japanese in Tokyo, Japan (JPT), Yoruba in Ibadan, Nigeria (YRI) and African ancestry in the Southwestern United States (ASW)) and our study so that comparisons to reference populations of known ethnicity could be made. A total of 2,546 ancestry-informative SNPs from the Illumina array were selected based on low inter-marker correlation and ability to differentiate between samples of African and European descent. An individual was subject to filtering from the analysis if his value along eigenvector 1 or 2 was outside of 4 standard deviations from the mean of each respective eigenvector. We identified 108 individuals who met this criterion. Eigenvector 1 was highly correlated ($\rho = 0.997$, $P < 1 \times 10^{-16}$) with percentage of European ancestry, estimated in HAPMIX[20]. Together, the top ten eigenvectors explain 21% of the global genetic variability among subjects.

Genotyping in the stage 2 and 3 studies was conducted using the TaqMan allelic discrimination assay. In stage 2, we removed samples missing data for greater than three SNPs ($n = 36$). To assess genotyping reproducibility, each study included replicate samples; the concordance was >98% for each SNP within each study. rs13116912 deviated from Hardy-Weinberg equilibrium in all but one of the stage 2 studies and was removed from the stage 2 analysis. No other SNP deviated from Hardy-Weinberg equilibrium ($P < 0.01$ in more than two studies) in stage 1 or 2. The call rate for rs7210100 was very high in stage 1 (99.9%) and was similar in cases (99.9%) and controls (99.9%). The call rate for this SNP was also very high in stages 2 (99.8% overall, 99.9% in cases and 99.8% in controls) and 3 (96.1% overall, 97.3% in cases and 95.5% in controls).

Sequencing. Bi-directional sequencing of rs7210100 and the five coding exons of *ZNF652* was performed in 48 subjects (20 homozygous for the risk variant, 20 heterozygous for the risk variant and 8 homozygous for the wild-type allele.) Primers were designed at least 50 bases upstream and downstream from each exon.

Statistical analysis. In stage 1, we tested the association of each SNP and prostate cancer risk using a 1-degree-of-freedom χ^2 likelihood ratio test from a logistic regression analysis adjusted for age, study and the first ten eigenvectors estimated by principal components analysis[19]. Overinflation of the test statistic was examined with and without adjustment for ancestry and was visualized with quantile-quantile plots. Lambdas were estimated as the median of the test statistics divided by 0.456 (the median of the 1-degree-of-freedom χ^2 null distribution). Age-adjusted ORs and 95% CIs for each SNP were estimated from the same logistic regression model. At each locus and for each participant, local ancestry was defined as the estimated number of European chromosomes (continuous between 0 and 2) carried by the participant estimated using the HAPMIX program[20]. Local ancestry at the 17q21 locus was evaluated as a confounder in the analysis of rs7210100.

Phased haplotype data from the founders of the CEU and YRI HapMap phase 2 samples were used to infer linkage disequilibrium patterns in order to impute untyped markers. We carried out genome-wide imputation using the software MACH[21]. The Rsq metric was used as a threshold in determining which SNPs to filter from analysis (Rsq < 0.3). Imputed SNPs in the 17q21 risk region, as shown in **Figure 2**, were examined in association with prostate cancer risk as described for typed SNPs above.

In stage 2, the SNPs were analyzed using logistic regression controlling for age and study (in the pooled analysis). Information regarding European ancestry was available for seven studies included in stages 2 and 3. As observed in stage 1 (**Supplementary Table 2**), the OR for rs7210100 was similar with and without adjustment for estimated European ancestry in these studies (**Supplementary Table 4**). The results for rs7210100 in stage 2, stage 3 and stages 1, 2 and 3 combined are presented without adjustment for ancestry. Heterogeneity of the OR across studies was evaluated using a likelihood ratio test.

Effect modification by age and first-degree family history of prostate cancer was assessed in stratified analyses, and significance was determined comparing the model with and without the cross-product term using a likelihood ratio test. We also examined the association of rs7210100 genotype with stage, Gleason score and the combination of stage and grade, with advanced disease defined as Gleason score ≥8 or stage ≥2 (non-localized disease), and non-advanced disease was defined as Gleason score <8 and stage = 1 (localized disease). A case-only analysis was used to test for differences in the association of rs7210100 with disease phenotypes. The association of rs7210100 with least-squares geometric-mean PSA levels was examined using multiple linear regression adjusting for age, body mass index and study.

We estimated the risk ratio between populations of different ancestral origin (African or European) caused by rs7210100 as $RR = [(1 - p_A)^2 + 2p_A(1 - p_A)RR_1 + p_A^2RR_2]/(1 - p_E)^2 + 2p_E(1 - p_E)RR_1 + p_E^2RR_2]$. Here p_A is the risk allele frequency in African origin populations, p_E is the risk allele frequency in European populations, RR_1 is the relative risk associated with carrying one copy of the risk allele (compared to none) and RR_2 is the relative risk associated with carrying two copies of the risk allele. We used values $p_A = 0.05$, $p_E = 0$, $RR_1 = 1.5$ and $RR_2 = 1.5^2$ so that the risk ratio between populations caused by the influence of this risk allele was estimated to be equal to 1.050625. Using the SEER incidence rates of prostate cancer

doi:10.1038/ng.839

in African Americans (234.6 per 100,000) and non-Hispanic individuals of European ancestry (150.4 cases per 100,000), we estimated the ratio of risks between these populations as 234.6/150.4 = 1.56. The percentage of greater risk to African Americans that may be associated with rs7210100 was estimated as $1 - [(1.56 - 1.050625)/(1.56 - 1)] \times 100$.

19. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
20. Price, A.L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
21. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).

ARTICLE

doi:10.1038/nature10336

The landscape of recombination in African Americans

A list of authors and their affiliations appears at the end of the paper

Recombination, together with mutation, gives rise to genetic variation in populations. Here we leverage the recent mixture of people of African and European ancestry in the Americas to build a genetic map measuring the probability of crossing over at each position in the genome, based on about 2.1 million crossovers in 30,000 unrelated African Americans. At intervals of more than three megabases it is nearly identical to a map built in Europeans. At finer scales it differs significantly, and we identify about 2,500 recombination hotspots that are active in people of West African ancestry but nearly inactive in Europeans. The probability of a crossover at these hotspots is almost fully controlled by the alleles an individual carries at *PRDM9* (P value $< 10^{-245}$). We identify a 17-base-pair DNA sequence motif that is enriched in these hotspots, and is an excellent match to the predicted binding target of *PRDM9* alleles common in West Africans and rare in Europeans. Sites of this motif are predicted to be risk loci for disease-causing genomic rearrangements in individuals carrying these alleles. More generally, this map provides a resource for research in human genetic variation and evolution.

In humans and many other species, recombination is not evenly distributed across the genome, but instead occurs in 'hotspots': 2-kilobase (kb) segments where the crossover rate is far higher than in the flanking DNA sequence[1–3]. The highest-resolution genetic map in contemporary humans so far—the deCODE map—is based on about 500,000 crossovers identified in 15,000 Icelandic meioses[4]. However, a limitation of maps built in people of European descent[4–6] is that they may not apply equally well in other populations, as suggested by comparisons of maps across ethnic groups[4,7–9] and patterns of linkage disequilibrium breakdown, which indicate that more of the genome may be recombinationally active in West Africans[10]. It is known that a major determinant of the positions of recombination hotspots is PRDM9, a meiosis-specific histone H3 methyltransferase whose zinc finger (ZF) domain binds DNA sequence motifs[11–13]. In Europeans, PRDM9 ZF arrays are predominantly of two similar types, A and B, both of which bind the 13-bp motif CCNCCNTNNCCNC[11]. In contrast, 36% of West African alleles are not of the A or B type[9,13]. Sperm typing of males who carry neither the A nor the B allele has shown no evidence of crossover activity at recombination hotspots associated with the 13-bp motif[9].

Building an African-American genetic map

To investigate differences in the crossover landscape across human populations, we built a genetic map in African Americans, who have an average of about 80% West African and 20% European ancestry, leading to genomes comprised of multi-megabase stretches of either West African or European ancestry[14]. Computational approaches, including HAPMIX[15], have been developed to infer the probability of 0, 1 or 2 European or African alleles at each locus in individuals genotyped at hundreds of thousands of single nucleotide polymorphisms (SNPs)[15–17]. Positions where the inferred number of European or African alleles changes reflect crossover events that have occurred since admixture began (on average six generations ago[15]). Change in the probability of European ancestry between adjacent SNPs can be interpreted as the probability of such a crossover between them. We inferred crossover events in 29,589 apparently unrelated African Americans who had been genotyped on SNP arrays in genetic association studies (Methods; Fig. 1a). To minimize false-positive crossovers, we restricted to crossovers that HAPMIX inferred with a probability of >95%, and that were flanked by a minimum of 2-centimorgan (cM) stretches where the ancestry was inferred to be unchanging (Supplementary Note 1). This produced 2,113,293 high-confidence crossovers, with a typical switch point resolved within 70 kb with probability 50% (Supplementary Note 1).

To build a high-resolution African-American genetic map (AA map), we leveraged the fact that most crossovers occur in hotspots shared across individuals[2] (Methods). Intuitively, although any crossover can only be roughly localized, inter-SNP intervals that are inferred to have an appreciable probability of crossover in multiple individuals are likely to contain recombination hotspots, allowing much better localization (Supplementary Fig. 1). To implement this idea, we modelled the recombination rate for each inter-SNP interval as shared across individuals and used Markov chain Monte Carlo (MCMC) to sample rates consistent with the data (Methods). This provides well-calibrated estimates of the crossing-over rate between all pairs of markers as well as estimates of rate uncertainty (Supplementary Note 1 and Supplementary Fig. 2). We find that the interval size at which the average recombination rate is equal to the standard error is 6 kb, which is the same accuracy that would be expected from a map based on 500,000 crossovers whose boundaries were precisely resolved (Supplementary Note 1). Despite this high resolution, there are also some limitations. First, the AA map does not separately infer male and female recombination rates (it is a sex-averaged map) and requires normalization by the total map length (like linkage disequilibrium maps[3,18]). Second, the map has less resolution and may miss a higher fraction of true crossovers at loci where it is more difficult to detect and resolve crossovers owing to low SNP density or low differentiation between West Africans and Europeans. Third, the map may be biased where ancestry deviates from the average, for example at chromosome 8q24, where the 10% of the people in this study who have prostate cancer have an increased proportion of African ancestry[19]. Fourth, the map assumes that all individuals are unrelated, whereas in fact there is probably some shared ancestry, resulting in multiple counting of some crossovers and an overestimation of map precision.

To assess the accuracy of the AA map, we generated an independent African-American pedigree map by analysing 222 nuclear families

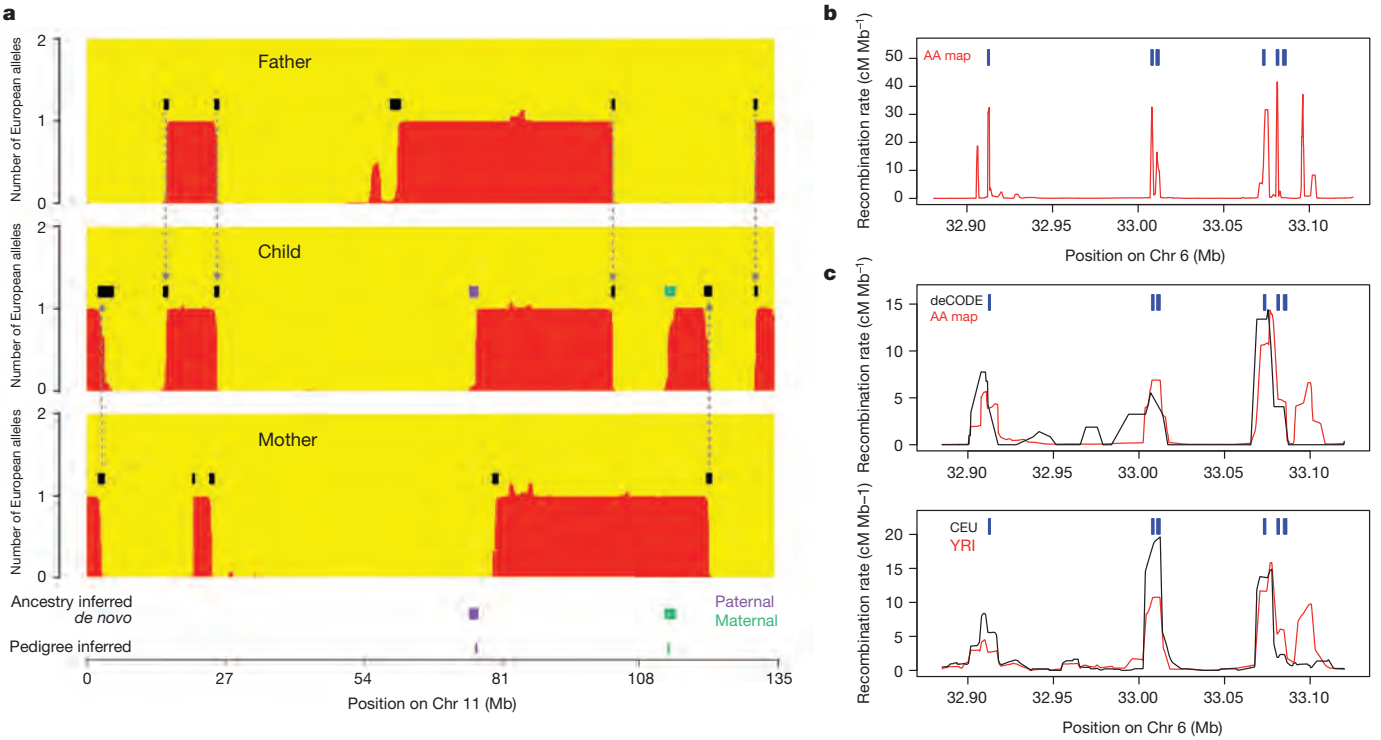


Figure 1 | Building an African-American genetic map. a, HAPMIX detection of crossovers between segments of inferred ancestry is illustrated in a father–mother–child trio. Black segments show inferred crossovers; arrows show transmission of ancestral crossovers from parent to child; purple/green segments show *de novo* events (paternal/maternal origin, respectively) corresponding to events identified directly using two additional children (bottom, 'pedigree inferred'). **b**, The AA map localizes five hotspots in a region of the MHC whose positions (blue) were previously mapped by sperm typing[1]. **c**, Comparison of maps shows a hotspot at 33.1 Mb in the African-derived AA and YRI maps, but not the deCODE and CEU maps (all maps smoothed to 10 kb).

that included 1,056 meioses in which we could directly detect crossovers between parent and child (Methods; Fig. 1a). Examination of the AA map rate around directly detected crossovers confirms the high resolution: the rate around such crossovers shows at least as strong a peak as that observed in maps based on linkage disequilibrium[2,3,18] (Supplementary Fig. 3). We next computed correlation coefficients for both the AA map and the deCODE map[4] to maps derived from the breakdown of linkage disequilibrium in Europeans (CEU) and West Africans (YRI)[18]. At broad scales (>3 Mb) they are almost identical ($\rho > 0.97$; Table 1). At fine scales, the AA map is more accurate (Table 1 and Supplementary Table 1), as reflected in a modest improvement in correlation to the CEU map at a 3-kb scale ($\rho_{AA,CEU} = 0.66$ versus $\rho_{deCODE,CEU} = 0.58$), and a major improvement for the YRI map, also at a 3-kb scale ($\rho_{AA,YRI} = 0.71$ versus $\rho_{deCODE,YRI} = 0.53$). The deCODE map is more correlated to the CEU map than to the YRI map at scales <1 Mb, suggesting that this map, built in Icelanders, reflects more European recombination rates. The AA map shows the opposite pattern, suggesting that it reflects more West African recombination patterns.

Population differences in hotspot locations

We compared the rate estimates for all four maps (AA, deCODE, CEU and YRI) over a 200-kb region within the major histocompatibility complex (MHC) locus where recombination rates in European males have been characterized through sperm typing[1] (Fig. 1b). The AA map detects five of six known hotspots, and localizes them to within 1 kb (the sixth hotspot is weak, with a peak male rate below the genome average[1]). Notably, the two maps based on samples with African ancestry (AA and YRI) found a hotspot not present in either map based on samples of European ancestry (deCODE and CEU) (Fig. 1c; Supplementary Fig. 4 gives a second example). We confirmed that such 'African-enriched' hotspots also occur genome-wide, by examining 2,375 loci with recombination rate peaks in the YRI map (>5 cM Mb^{-1}) but not the CEU map (<1 cM Mb^{-1}), and finding a rate rise in the independently generated AA map, but not in the deCODE map (Supplementary Fig. 5A). In the reciprocal experiment searching for European-specific hotspots, we find no such evidence for genuine ancestry specificity; at loci with recombination rate peaks in the CEU map but not the YRI map, there are weak peaks in both the deCODE and AA maps

Table 1 | Genetic map assessments at different size scales

Scale (interval size)	Pearson correlation (ρ) of the AA map (deCODE map) to the specified LD map			Estimated correlation of AA map to the true map (inferred by MCMC)†	Estimated coefficient of variation of AA map (s.e. divided by crossover rate expected for interval size)†
	Combined LD*	CEU	YRI		
3 kb	0.75 (0.63)	0.66 (0.58)	0.71 (0.53)	0.93	1.41
10 kb	0.82 (0.74)	0.73 (0.70)	0.78 (0.65)	0.96	0.73
30 kb	0.86 (0.83)	0.78 (0.78)	0.83 (0.74)	0.98	0.36
100 kb	0.91 (0.89)	0.84 (0.85)	0.87 (0.81)	0.99	0.17
300 kb	0.94 (0.93)	0.89 (0.90)	0.92 (0.88)	1.00	0.08
1 Mb	0.97 (0.96)	0.94 (0.94)	0.95 (0.95)	1.00	0.04
3 Mb	0.98 (0.98)	0.97 (0.97)	0.98 (0.97)	1.00	0.02

The numbers in this table are restricted to the autosomes and genomic segments more than 5 Mb from the telomeres. LD, linkage disequilibrium; s.e., standard error.
* The combined map is the HapMap2 population-averaged linkage-disequilibrium-based map[18].
† The s.e. of the map at each size scale is determined by the posterior probability distribution from the MCMC.

(Methods and Supplementary Fig. 5B). Thus, hotspots active in Europeans are consistently 'shared' with YRI and African Americans, whereas populations with African ancestry harbour additional, non-shared hotspots that we call 'African-enriched'.

Mapping variants underlying population differences

To understand the features of recombination in West Africans that differ from Europeans, we estimated the degree to which each African-American person's crossovers occur in African-enriched hotspots, compared with shared hotpots, a phenotype we refer to as their African enrichment (AE). We view each individual's crossovers as sampled from a mixture of two genetic maps—an 'S map' of shared hotspots based on the deCODE map, and an 'AE map' of African-enriched hotspots that is learned from comparing the deCODE and AA maps—so that the proportion of crossovers assigned to the AE map is a person's AE phenotype (Supplementary Note 4). We tested approximately 3 million SNPs (genotyped and imputed) for association with three phenotypes: AE, usage of linkage-disequilibrium-based hotspots known to be enriched for the 13-bp motif CCNCCNTNNCCNC[20] and genome-wide crossover rate (in pedigrees) (Methods and Supplementary Note 4). In crossovers detected in unrelated African Americans, the alleles a person carries are only sometimes descended from the ancestor in whom the crossover occurred, thus adding noise to the association signal (nevertheless there is useful signal given the large sample size; Supplementary Note 4). In the pedigree map, association between alleles and AE can be tested directly because we have genotypes in the parents.

The SNP showing the strongest association with AE is rs6889665 ($P = 1.5 \times 10^{-246}$; Fig. 2a and Supplementary Fig. 6), which has a derived allele frequency of 29% in YRI and 2% in CEU, and is within 4 kb of the ZF array of *PRDM9* (refs 4, 9, 11–13). This SNP is associated with AE in both the pedigree individuals and the unrelated individuals (Supplementary Note 4), and is also the SNP most strongly associated with usage of linkage-disequilibrium-based hotspots ($P = 1.8 \times 10^{-52}$) (Supplementary Table 2). No locus outside *PRDM9* is significant ($P < 0.01$ after Bonferroni correction; Supplementary Table 2). To understand better the association at rs6889665, we inferred the alleles in the *PRDM9* ZF array carried

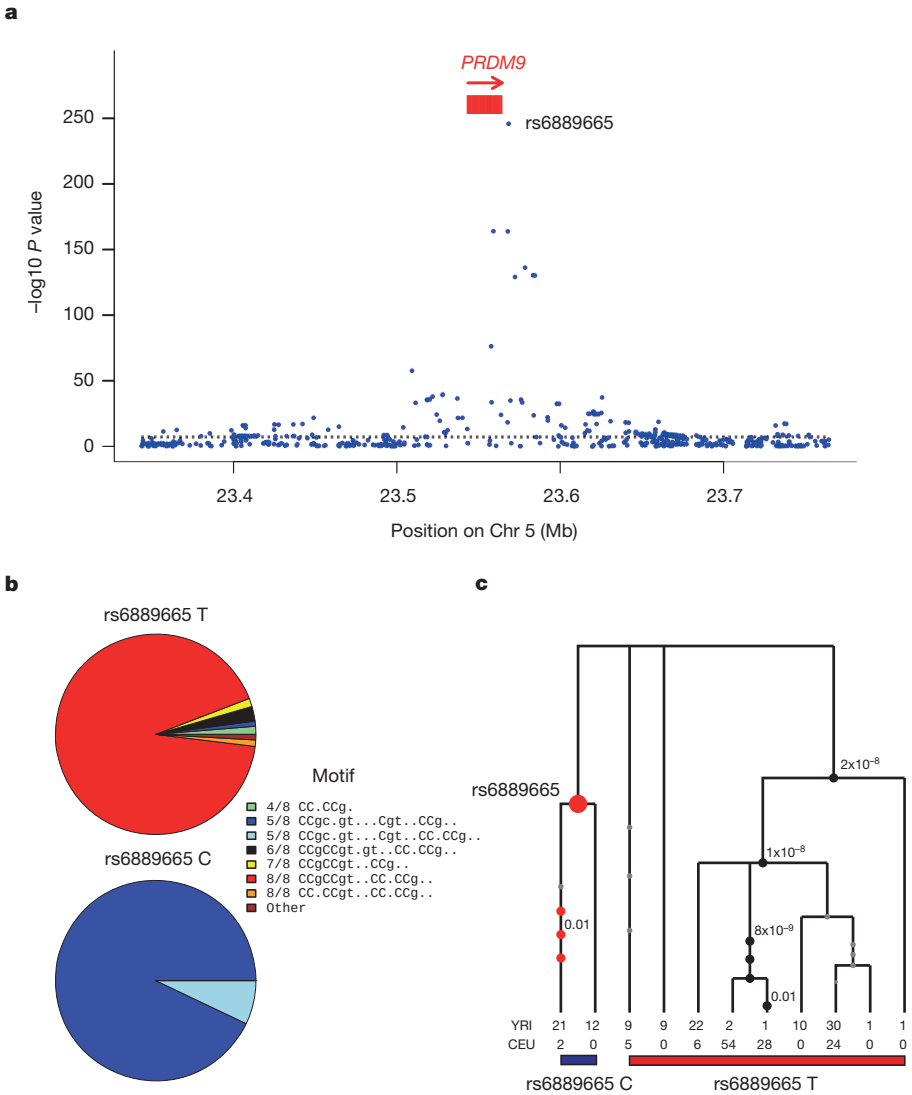


Figure 2 | Association of *PRDM9* genetic variation with hotspot activity. **a**, A genome-wide association study measuring association of the AE phenotype shows a single genome-wide significant peak at *PRDM9*, with rs6889665 the best-associated SNP. **b**, Relationship between alleles of rs6889665 and predicted binding target of the *PRDM9* ZF array[9] for West African and European samples. The binding predictions are grouped into 8 clusters according to their best-matching region to the 13-bp motif, and annotated by the number of bases matching the motif. The African-enriched rs6889665 C allele always co-occurs with motifs with a poor (5/8) match to the 13-bp motif. **c**, Gene tree[25] of the linkage disequilibrium block containing the *PRDM9* ZF array (Methods); numbered circles show SNPs and significant *P* values for association, after conditioning on rs6889665.

by 139 individuals based on sequencing data from the 1000 Genomes Project[10], using the reads to infer each individual's *PRDM9* alleles among 29 alleles whose full sequences were previously determined[9] (Supplementary Note 5). Grouping *PRDM9* alleles on the basis of how closely their binding target predictions match the 8 non-degenerate bases of the 13-bp motif, following a previously described approach[9], we find that the ancestral 'T' variant at rs6889665 is strongly correlated to alleles with an exact (8/8) match to the 13-bp motif (including the A and B alleles), whereas the derived 'C' variant is almost perfectly correlated to a group of alleles, all predicted to bind a common, different 17-bp motif—CCgCNgtNNNCgtNNCC[9]—which matches the 13-bp motif at only 5 bases (5/8 match; less strongly signalled bases in the motif are in lowercase and 'N' may be any base). This implies a common historical origin for alleles matching this 17-bp motif (Fig. 2b, Supplementary Fig. 7 and Supplementary Note 5). We also experimentally measured the number of ZF domains in *PRDM9* in 354 individuals including 166 African Americans from the pedigree study (Methods). This showed, again, that rs6889665 differentiates *PRDM9* alleles into two different classes, with 96% of haplotypes carrying the ancestral allele having <14 ZFs, and 93% of haplotypes carrying the derived allele having ≥14 ZFs (Supplementary Fig. 7). After conditioning on rs6889665, there is no evidence that ZF array length is associated with the AE phenotype. Several SNPs near the *PRDM9* ZF array show a conditional association signal that is much weaker than rs6889665, but still significant (Fig. 2c, Supplementary Fig. 6 and Supplementary Note 4), with the strongest at rs10043097 ($P = 8.3 \times 10^{-14}$), upstream of the *PRDM9* transcription start site. These SNPs may tag additional variation in the *PRDM9* ZF array, or potentially expression levels.

Finding a motif for African-enriched hotspots

To identify directly candidate African-enriched hotspot motifs, we selected 2,454 loci with a high crossover rate in the AE map and YRI map (>2 cM Mb^{-1} over 2 kb), and no more than half this rate in the S map and CEU map (this set is more powerfully enriched for higher recombination in people of African ancestry than the 2,375 above, as it includes information from the contemporary maps). We compared these to a 'control set' of 7,328 candidate hotspots more active in the European- than the African-derived maps (Methods and Supplementary Note 6). To identify sequence motifs associated with the African-enriched hotspots[3,21], we identified short motifs that occurred at increased frequency in the African-enriched hotspot set (Supplementary Note 6). Testing all motifs with lengths of 5–9 bases revealed a 9-nucleotide motif CCCCAGTGA (odds ratio (OR) = 1.79, $P = 2.24 \times 10^{-8}$, Bonferroni corrected $P = 0.004$), which exhibited a kilobase-scale rate peak near occurrences of this motif in African-derived maps, but in neither of the European-derived maps (Supplementary Fig. 8). Further analysis revealed a strong influence of downstream flanking bases (Supplementary Fig. 9) and degeneracy, yielding a 17-bp consensus sequence, CCCCaGTGAGCGTtgCc (Fig. 3a; more strongly signalled bases are in uppercase), with the same consensus obtained when we considered flanking sequences for only odd or even chromosomes, and whether we based the analysis on AE–S or YRI–CEU map comparisons (Supplementary Note 6). The 500 best matches to this motif have a ~3-fold increase in average rate in the AA and YRI relative to the deCODE and CEU maps (Fig. 3b and Supplementary Fig. 8G). Hotspots associated with the motif occur in both unique and repetitive DNA (for example, L1PA10/13 LINE elements; Supplementary Fig. 10 and Supplementary Note 6). We also compared the 17-bp consensus to the binding motif predicted for 5/8 match alleles, and found that they match almost precisely (Fig. 3a; 10 of 11 bases, $P = 8.1 \times 10^{-6}$).

Assessing the impact of *PRDM9* on recombination

How much of the African-enriched recombination pattern can be explained by *PRDM9*? We estimated the fraction of variation in the AE phenotype explained by rs6889665 in our pedigree data after accounting for noise in the phenotype estimation (Supplementary Note 4). Over 82% of map usage variability is explained by the rs6889665 genotype alone. Given that there are further influential *PRDM9* variants (Fig. 2c), this gene may thus explain almost all differences in local rate between the West African and European populations. We next examined rates around 82 narrowly defined (<10 kb) crossover sites in 7 individuals homozygous for the derived allele at rs6889665. There is no evidence of hotspots at these loci in either the deCODE or CEU maps (Fig. 3c), in contrast to crossovers in individuals carrying the ancestral allele at rs6889665 (Supplementary Fig. 11). Thus, crossover positions in individuals who are homozygous for the derived allele at rs6889665 are consistent with an entirely different recombination hotspot landscape, which would imply *PRDM9* control of all hotspots[9]. Despite the strong correlation between maps at megabase scales, there is mounting evidence that *PRDM9*'s influence

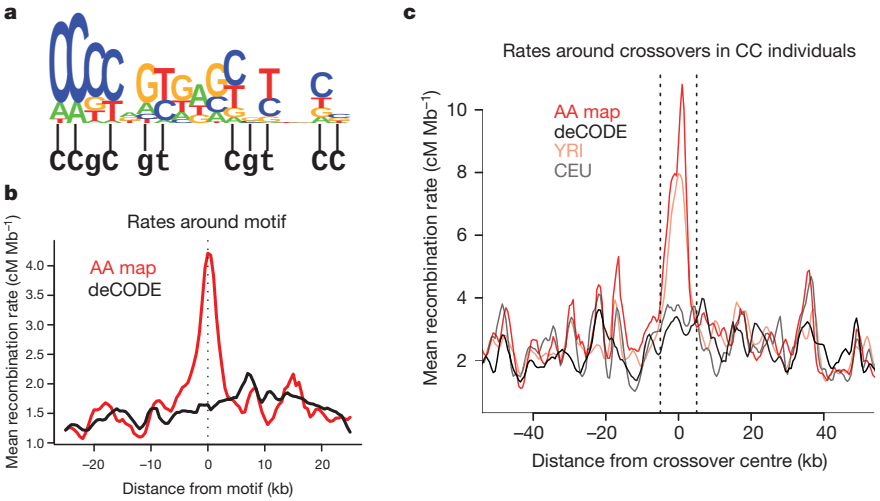


Figure 3 | A sequence motif specifying the positions of African-enriched hotspots. **a**, Logo plot showing a degenerate 17-bp hotspot motif, with stack height proportional to −log *P* value, and relative letter height proportional to the mean crossover rate increase given each base. Below is the bioinformatic *PRDM9* binding prediction for the alleles associated with rs6889665 allele C (from Fig. 2b), matching this motif at 10/11 bases (lines). **b**, Average crossover rate (in 2-kb sliding windows) in the AA (red line) and deCODE (black line) maps surrounding the 500 strongest motif matches. **c**, In seven rs6889665 CC individuals from the pedigree study, we localized 82 crossovers to within 10 kb, and plot average AA, YRI, deCODE and CEU map rates. There is no strong peak above local background in the deCODE or CEU maps.

on crossing over may not be limited to fine scales[4,11]: we observe a weakly significant association of rs6889665 with the total number of crossovers genome-wide in pedigrees ($P = 0.04$), corresponding to an average 1.3 crossovers more per meiosis per derived allele, exceeding the strongest previously known association[22] at *RNF212*.

Conclusions

We have shown that *PRDM9* alleles that bind a novel 17-bp motif and occur at greatly increased frequency in people of West African ancestry have led to a shift in the recombination landscape compared with people of non-African ancestry. The larger number of hotspots available to West Africans implies that at the population level, crossovers are more evenly distributed than in Europeans[10], and thus the shorter extent of West African linkage disequilibrium is not due to differences in demographic history alone (such as the lack of an out-of-Africa founder event)[23]. Our findings also have medical implications, as recombination errors leading to insertions or deletions are known to be associated with recombination hotspots[9,21,24]. Our results predict that the congenital abnormalities that have been associated with the recombination hotpots bound by *PRDM9* A and B alleles will occur at a decreased rate in people of West African ancestry, whereas new diseases will arise due to recombination errors near African-enriched hotspots.

METHODS SUMMARY

We assembled SNP array data from 29,589 unrelated people and 222 nuclear families genotyped at 490,000–910,000 SNPs from the Candidate Gene Association Resource (CARe), studies at the Children's Hospital of Philadelphia (CHOP), the African American Breast Cancer Consortium, the African American Prostate Cancer Consortium and the African American Lung Cancer Consortium. To build a recombination map, we used HAPMIX to localize candidate crossover positions[15], and implemented a MCMC that used the probability distributions for the positions of the filtered crossovers to infer recombination rates for each of 1.3 million inter-SNP intervals. We also implemented a second MCMC that models each individual's set of crossovers as a mixture of an S map, similar to the European deCODE map, and an AE map, and then assigned each individual an 'AE phenotype' corresponding to the proportion of their newly detected crossovers assigned to the AE map. We imputed genotypes at up to three million HapMap2 SNPs[18] and then tested each of these SNPs for association with the AE phenotype and other recombination-related phenotypes. We identified 2,454 candidate African-enriched hotspots with increased recombination rates in the YRI versus CEU maps, and in the AE versus S maps, and searched for motifs enriched at these loci, thus identifying a degenerate 17-bp motif. To study the structure of *PRDM9*, we measured the length of the *PRDM9* ZF array and genotyped rs6889665 in YRI, CEU and the CARe nuclear families; we also carried out imputation based on 1000 Genomes Project short read data[10] to infer the alleles individuals carry, among 29 previously characterized in a sequencing study of *PRDM9* (ref. 9).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 February; accepted 27 June 2011.
Published online 20 July 2011.

1. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29,** 217–222 (2001).
2. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304,** 581–584 (2004).
3. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310,** 321–324 (2005).
4. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467,** 1099–1103 (2010).
5. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31,** 241–247 (2002).
6. Matise, T. C. *et al.* A second-generation combined linkage-physical map of the human genome. *Genome Res.* **17,** 1783–1786 (2007).
7. Weitkamp, L. R. Proceedings: population differences in meiotic recombination frequency between loci on chromosome 1. *Cytogenet. Cell Genet.* **13,** 179–182 (1974).
8. Jorgenson, E. *et al.* Ethnicity and human genetic linkage maps. *Am. J. Hum. Genet.* **76,** 276–290 (2005).
9. Berg, I. L. *et al. PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genet.* **42,** 859–863 (2010).
10. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
11. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327,** 836–840 (2010).
12. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327,** 876–879 (2010).
13. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327,** 835 (2010).
14. Smith, M. W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74,** 1001–1013 (2004).
15. Price, A. L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5,** e1000519 (2009).
16. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* **82,** 290–303 (2008).
17. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74,** 979–1000 (2004).
18. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
19. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103,** 14068–14073 (2006).
20. Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319,** 1395–1398 (2008).
21. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hotspots and genome instability in humans. *Nature Genet.* **40,** 1124–1129 (2008).
22. Kong, A. *et al.* Sequence variants in the *RNF212* gene associate with genome-wide recombination rate. *Science* **319,** 1398–1401 (2008).
23. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411,** 199–204 (2001).
24. Raedt, T. D. *et al.* Conservation of hotspots for recombination in low-copy repeats associated with the *NF1* microdeletion. *Nature Genet.* **38,** 1419–1423 (2006).
25. Griffiths, R. C. & Tavaré, S. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.* **127,** 77–98 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the participants who donated DNA samples, to D. Altshuler, J. Buard, K. Bryc, J. Kovacs, B. de Massy, G. McVean, B. Pasaniuc and S. Sankararaman for conversations and critiques, and to A. Auton for facilitating analysis of the 1000 Genomes Project data. Analysis was supported by the Wellcome Trust and NIH grants HL084107 and GM091332. CARe was supported by a contract from the National Heart, Lung and Blood Institute (HHSN268200960009C) to create a phenotype and genotype database for dissemination to the biomedical research community. Eight parent studies contributed phenotypic data and DNA samples through the Broad Institute (N01-HC-65226): the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA) study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS) and the Sleep Heart Health Study (SHHS). Support for CARe also came from the individual research institutions, investigators, field staff and study participants. Individual funding information is available at http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx. All genome-wide genotyping of samples from the Children's Hospital of Pennsylvania (CHOP) was supported by an Institutional Development Award to the Center for Applied Genomics from the Children's Hospital of Philadelphia, a research award from the Landenberger Foundation and the Cotswold Foundation. We thank all study participants and the staff at the Center for Applied Genomics for performing the genotyping. The African American Breast Cancer Consortium (AABCC) was supported by a DoD Breast Cancer Research Program Era of Hope Scholar Award to C.A.H. and the Norris Foundation, and by grants to the component studies: MEC (CA63464, CA54281); CARE (HD33175); WCHS (CA100598, DAMD 170100334, Breast Cancer Research Foundation); SFBC (CA77305, DAMD 17966071); CBCS (CA58223, ES10126), PLCO (NCI Intramural Research Program); NHBS (CA100374); WFBC (R01-CA73629); and CPS-II (the American Cancer Society). The African American Prostate Cancer Consortium (AAPCC) was supported by grants CA63464, CA54281, CA1326792, CA148085 and HG004726, and by grants to the component studies: PLCO (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), both MEC and LAAPC (PC35139, DP000807); MDA (CA68578, CA140388, ES007784, DAMD W81XWH0710645); GECAP (ES011126); CaP Genes (CA88164); IPCG (W81XWH0710122); DCPC (GM08016, DAMD W81XWH0710203, DAMD W81XWH0610066); and SCCS (CA092447, CA68485). The African American Lung Cancer Consortium (AALCC) was supported by grants CA060691, CA87895, PC35145 and CA22453, CA68578, CA140388, ES007784, ES06717, CA55769, CA127219, CA1116460S1, CA1116460, CA121197, CA141716, CA121197S2, CPRIT RP100443, CA148127, DAMD W81XWH0710645, University Cancer Foundation, Duncan Family institute, Center for Community, Implementation and Dissemination Research Core, and by grants to the component studies: PLCO and the Maryland Studies (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), and both MEC and LAAPC (PC35139, DP000807).

Author Contributions D.R. and S.R.M. conceived the study. A.G.H., A.T., N.P., Y.S., N.R., C.D.P., G.K.C., K.W., S.G.B., D.R. and S.R.M. performed analyses. N.R. performed the experimental work (genotyping of polymorphisms at *PRDM9*). A.G.H., N.P., J.N.H.,

B.E.H., H.A.T. Jr, A.L.P., H.H., S.J.C., C.A.H., J.G.W., D.R. and S.R.M. coordinated the study. A.G.H., D.R. and S.R.M. wrote the paper. N.R., C.D.P., G.K.C., K.W., S.G.B., S.R., J.N.H., B.E.H., H.A.T. Jr, H.H., S.J.C., C.A.H., J.G.W., D.R. and all the alphabetically listed authors contributed to sample collection and generation of SNP array data. All authors contributed to revision and review of the manuscript.

Author Information Crossover rate estimates for the AA map can be found at http://www.well.ox.ac.uk/~anjali/AAmap/. We also provide estimates of uncertainty for the map based on samples from the MCMC. Association testing results for each SNP are available from the authors on request. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or S.R.M. (myers@stats.ox.ac.uk).

Anjali G. Hinch[1], Arti Tandon[2,3], Nick Patterson[2], Yunli Song[4], Nadin Rohland[2,3], Cameron D. Palmer[5,6], Gary K. Chen[7], Kai Wang[8,9], Sarah G. Buxbaum[10], Ermeg L. Akylbekova[10,11], Melinda C. Aldrich[12,13], Christine B. Ambrosone[14], Christopher Amos[15], Elisa V. Bandera[16], Sonja I. Berndt[17], Leslie Bernstein[18], William J. Blot[13,19], Cathryn H. Bock[20], Eric Boerwinkle[21], Qiuyin Cai[13], Neil Caporaso[17], Graham Casey[7], L. Adrienne Cupples[22], Sandra L. Deming[13], W. Ryan Diver[23], Jasmin Divers[24], Myriam Fornage[25], Elizabeth M. Gillanders[26], Joseph Glessner[9], Curtis C. Harris[27], Jennifer J. Hu[28], Sue A. Ingles[7], William Isaacs[29], Esther M. John[30], W. H. Linda Kao[31], Brendan Keating[9], Rick A. Kittles[32], Laurence N. Kolonel[33], Emma Larkin[34], Loic Le Marchand[33], Lorna H. McNeill[35], Robert C. Millikan[36], Adam Murphy[37], Solomon Musani[11], Christine Neslund-Dudas[38], Sarah Nyante[36], George J. Papanicolaou[39], Michael F. Press[7], Bruce M. Psaty[40], Alex P. Reiner[41], Stephen S. Rich[42], Jorge L. Rodriguez-Gil[28], Jerome I. Rotter[43], Benjamin A. Rybicki[38], Ann G. Schwartz[20], Lisa B. Signorello[13,19], Margaret Spitz[15], Sara S. Strom[44], Michael J. Thun[23], Margaret A. Tucker[17], Zhaoming Wang[45], John K. Wiencke[46], John S. Witte[47], Margaret Wrensch[46], Xifeng Wu[15], Yuko Yamamura[44], Krista A. Zanetti[26,27], Wei Zheng[13], Regina G. Ziegler[17], Xiaofeng Zhu[48], Susan Redline[49], Joel N. Hirschhorn[5,6,50], Brian E. Henderson[7], Herman A. Taylor Jr[11,51,52], Alkes L. Price[53], Hakon Hakonarson[9,54], Stephen J. Chanock[17],Christopher A. Haiman[7], James G. Wilson[55], David Reich[2,3]* & Simon R. Myers[1,4]*

[1]Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK. [2]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Deptartment of Genetics, Harvard Medical School, New Research Building, 77 Ave. Louis Pasteur, Boston, Massachusetts 02115, USA. [4]Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK. [5]Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [6]Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital Boston, Massachusetts 02115, USA. [7]Department of Preventive Medicine and Department of Pathology, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California 90033, USA. [8]Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, California 90089, USA. [9]Center for Applied Genomics, The Childrens Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [10]Jackson Heart Study Coordinating Center, Jackson State University, 350 W. Woodrow Wilson Ave., Suite 701, Jackson, Mississippi 39213, USA. [11]Department of Medicine, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA. [12]Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. [13]Division of Epidemiology in the Department of Medicine, Vanderbilt Epidemiology Center; and the Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. [14]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York 14263, USA. [15]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [16]The Cancer Institute of New Jersey, New Brunswick, New Jersey 08903, USA. [17]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. [18]Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, California 91010, USA. [19]International Epidemiology Institute, Rockville, Maryland 20850, USA. [20]Karmanos Cancer Institute and Department of Oncology, Wayne State University of Medicine, Detroit, Michigan 48201, USA. [21]Human Genetics Center and Division of Epidemiology, University of Texas at Houston, 1200 Herman Pressler St., Houston, Texas 77030, USA. [22]Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, Massachusetts 02118 and Framingham Heart Study, Framingham, Massachusetts 01702, USA. [23]Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. [24]Department of Biostatistical Sciences, Wake Forest University School of Medicine WC-2326, Medical Center Blvd., Winston Salem, North Carolina 27157, USA. [25]Institute of Molecular Medicine and Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center at Houston, 1825 Pressler Street, Houston, Texas 77030, USA. [26]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland 20892, USA. [27]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA. [28]Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. [29]James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institutions, Baltimore, Maryland 21287, USA. [30]Cancer Prevention Institute of California, Fremont, California 94538; and Stanford University School of Medicine and Stanford Cancer Center, Stanford, California 94305, USA. [31]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, Maryland 21205, USA. [32]Department of Medicine, University of Illinois at Chicago, Chicago, Illinois 60607, USA. [33]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. [34]Department of Medicine, Division of Allergy, Pulmonary and Critical Care, 6100 Medical Center East, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8300, USA. [35]Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, and Center for Community Implementation and Dissemination Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. [36]Department of Epidemiology, Gillings School of Global Public Health, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [37]Department of Urology, Northwestern University, Chicago, Illinois 60611, USA. [38]Department of Public Health Sciences, Henry Ford Hospital, Detroit, Michigan 48202, USA. [39]Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, 6701 Rockledge Drive, Bethesda, Maryland 20892, USA. [40]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology & Health Services, University of Washington; Group Health Research Institute; Group Health Cooperative; 1730 Minor Ave., Seattle, Washington 98101, USA. [41]Department of Epidemiology, University of Washington, Box 357236 Seattle, Washington 98195, USA. [42]Center for Public Health Genomics, University of Virginia, West Complex Room 6111, Charlottesville, Virginia 22908, USA. [43]Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. [44]Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. [45]Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland 20877, USA. [46]University of California San Francisco, San Francisco, California 94158, USA. [47]Institute for Human Genetics, Departments of Epidemiology and Biostatistics and Urology, University of California, San Francisco, San Francisco, California 94158, USA. [48]Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Wolstein Research Building, Cleveland, Ohio 44106, USA. [49]Brigham and Women's Hospital, Department of Medicine, Division of Sleep Medicine, 75 Francis Street, Boston, Massachusetts 02115, USA. [50]Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [51]Jackson State University, 1400 Lynch Street, Jackson, Mississippi 39217, USA. [52]Tougaloo College, 500 West County Line Road, Tougaloo, Mississippi 39174, USA. [53]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [54]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [55]Department of Physiology and Biophysics, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA.
*These authors contributed equally to this work.

METHODS

Samples used for building the AA map. The 29,589 unrelated African-American samples derive from five sources. Informed consent was provided by all the individuals participating in the study, and was approved by all of the institutions responsible for sample collection.

The first source is the Candidate Gene Association Resource (CARe) study, a consortium of cohorts. We analysed CARe samples genotyped on the Affymetrix 6.0 array from the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). After removing individuals known to be related, and restricting to SNPs with good completeness in all cohorts, we had data from 6,209 individuals typed at 580,000 SNPs.

The second source consists of diverse studies carried out at the Children's Hospital of Philadelphia (CHOP), which has established a biobank for Philadelphia children to facilitate large genotype–phenotype association analysis. The cohort was recruited by CHOP clinicians, nursing and medical assistant staff within the CHOP Health Care Network, including primary care clinics and outpatient practices, from the hospital's patient base of over one million paediatric patients. All samples analysed here were genotyped on either the Illumina 610-Quad or Illumina HumanHap550 array. After removing individuals known to be related, identifying American Americans by multidimensional scaling on genotype data, and restricting to SNPs with a high level of completeness across samples, we had data from 7,503 samples typed at 491,572 SNPs.

The third source is the African American Breast Cancer Consortium (AABCC), consisting of the Multiethnic Cohort study (MEC), the Los Angeles component of the Women's Contraceptive and Reproductive Experiences study (CARE), the Women's Circle of Health Study (WCHS), the San Francisco Bay Area Breast Cancer study (SFBC), the Carolina Breast Cancer Study (CBCS), the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Cohort (PLCO), the Nashville Breast Health Study (NBHS) and the Wake Forest University Breast Cancer Study (WFBC), all genotyped on an Illumina 1M array. After data curation, including removal of samples with genetic evidence of being second-degree relatives or closer using the *smartrel* package of EIGENSOFT[26] (>0.2 correlation of genotype state), we had data from 5,203 women (about half cases and half controls) typed at 894,717 SNPs.

The fourth source is the African American Prostate Cancer Consortium (AAPCC), consisting of the MEC, the Southern Community Cohort Study (SCCS), PLCO, the Cancer Prevention Study II Nutrition Cohort (CPS-II), the Prostate Cancer Case-Control Studies at MD Anderson (MDA), the Identifying Prostate Cancer Genes study (IPCG), the Los Angeles Study of Aggressive Prostate Cancer (LAAPC), the Prostate Cancer Genetics Study (CaP Genes), the Case-Control Study of Prostate Cancer among African Americans in Washington DC (DCPC), the Gene-Environment Interaction in Prostate Cancer Study (GECAP) and the Cancer Prevention Study II (CPS-II), all typed on an Illumina 1M array. After the same data curation as the breast cancer study, we had data from 6,540 men (about half cases and half controls) typed at 896,036 SNPs.

The fifth source is individuals from the African American Lung Cancer Consortium (AALCC), including cases and controls from the MEC, the SCCS, PLCO, the MD Anderson (MDA) African American Lung Cancer Study, the NCI-Maryland Lung Cancer Case-Control Study, the University of California at San Francisco African American Lung Cancer Study and the Wayne State African American Lung Cancer Study, all genotyped on the Illumina 1M array. After data curation, we had data from 4,134 individuals typed at 906,687 SNPs.

Samples used for building the pedigree map. The pedigree map was built using data from 135 African-American nuclear families from CARe and 87 African-American families from CHOP for which genotyping data were available from at least two full siblings and at least one parent. The CARe studies that contributed samples were JHS (70 families, including 58 samples that we newly genotyped on the Affymetrix 6.0 array to increase the number of crossovers we could analyse) and CFS (65 families). For the families with a missing parent, we developed a Hidden Markov Model (HMM) approach to jointly estimate the genotype of the missing parent as well as to infer the position of crossover events in the offspring. The observed variables in the HMM were the genotypes of the available family members and the states of the HMM were the genotypes of the parents and the identity by descent (IBD) status of the children. A change in IBD status in an offspring is interpreted as a crossover event. Supplementary Note 2 provides details of the HMM used to infer positions of these pedigree crossover events.

Local ancestry inference and identification of crossover events. We merged the data for each cohort with phased YRI and CEU data from the HapMap3 data set[27]. We filtered SNPs that had a frequency inconsistent with an 80–20% linear combination of YRI and CEU frequencies (*t* statistic with an absolute value of greater than 3), potentially reflecting genotyping error in either the HapMap3 or the cohort data.

We ran HAPMIX on these data using a prior hypothesis of 20% European ancestry and 6 generations since mixture for each individual[15]. HAPMIX requires users to input a recombination map as a prior distribution, and we assumed that rates were constant across each chromosome arm with a total rate across each arm determined by the Rutgers genetic map[6] (Supplementary Note 1).

Filtering of crossover events had three stages. First, we removed crossover events where the probability of occurrence was estimated to be less than 95% by HAPMIX. Second, we removed candidate crossover events that were non-monotonic, that is, where the probability of an overlapping crossover event with an ancestry switch in a different direction was ≥1% within any inter-SNP interval. Third, we removed crossover events where either of the two flanking ancestry blocks was smaller than 2 cM in size as measured with respect to a published map based on linkage disequilibrium[3,18] (Supplementary Note 1). For comparisons to the deCODE map and linkage-disequilibrium-based maps, we also removed segments of the genome within 5 Mb of the telomeres (to be consistent with the comparisons presented in the deCODE study where the same restriction was applied[4]).

Construction of the AA map. All 22 autosomes and chromosome X were split into approximately 1.3 million inter-SNP intervals based on the union of SNPs analysed across all five sample sets. Our goal was to estimate a crossover rate for each of these intervals. We modelled crossover rates such that the rate for each SNP interval is independent of every other SNP interval, motivated by a hotspot model. We used a gamma prior on rates with the mean estimated from the filtered HAPMIX output (Supplementary Note 1). We used a Gibbs sampler to sample rates in every SNP interval and to determine the location of a crossover event within the 95% range estimated by the HAPMIX output. In each round of the Gibbs sampler, we used the set of sampled rates in the previous round to construct a probability mass function for the SNP interval in which each crossover occurred, using an approach described in Supplementary Note 1 to approximate the probability mass function that HAPMIX would have produced conditional on the previous set of sampled rates. After sampling the location of the crossover events, we counted how many crossovers occurred in every SNP interval. We used these counts to construct a posterior distribution for the crossover rate in each SNP interval, taking advantage of the conjugacy of a Poisson likelihood and a gamma prior. We then sampled a crossover rate for each SNP interval from its respective gamma posterior distribution.

Candidate African-enriched hotspots. To identify candidate African-enriched hotspots, we used two pairs of maps: the previously available YRI map and CEU map, and the AE map and the S map. We combined information from both map pairs to enrich for regions with genuine differences between the West African and European populations. Specifically, we identified candidate hotspots as 2-kb intervals representing a peak in the AE map rate, where the estimated rate in the AE map was $>2\,\mathrm{cM\,Mb^{-1}}$ and at least double that in the S map, and in addition the YRI map rate was $>2\,\mathrm{cM\,Mb^{-1}}$ and at least double the CEU map rate. We took the resulting candidate hotspot set and defined hotspot boundaries by identifying the region flanking the 2 kb rate peak that had rates at least 50% of the peak value in the AE map. Regions larger than 5 kb were discarded. We similarly constructed a set of 'shared' hotspots but modified the initial criteria given the lack of obvious hotspots present only in people of European ancestry. Specifically, we identified 2 kb S map rate peak locations where both the S and CEU estimated rates were $>2\,\mathrm{cM\,Mb^{-1}}$, while the AE and YRI map rates were below those in these respective European populations. We then narrowed the regions and filtered using the same procedure we had developed for the candidate African-enriched hotspots.

Association testing. MaCH[28] was used to impute up to 3,058,149 SNP genotypes from HapMap2 (ref. 18) into all African Americans we analysed, using the unrelated YRI and CEU samples as combined reference panels. We tested for association at all SNPs with minor allele frequency > 1%. To restrict our analysis to individuals in whom the phenotype was measured accurately, we performed the association analysis with the AE and hotspot usage phenotypes only in individuals with at least 35 inferred crossovers. Association testing was carried out using linear regression, after controlling for gender, genome-wide European ancestry proportion (inferred by HAPMIX) and study (Supplementary Note 4). We observe slight inflation of the association statistics genome-wide compared with the expectation (the Genomic Control inflation factor[29] is 1.046 for the AE phenotype and 1.038 for the hotspot usage phenotype), which we propose may reflect cryptic relatedness among samples (Supplementary Note 4). We report *P* values after correction using Genomic Control[29].

Construction of *PRDM9* tree. To examine the history of the *PRDM9* ZF array and to place SNPs showing association with AE map usage within the framework of this history, we identified 19 SNPs from HapMap2 (ref. 18) that surrounded the

ZF array and that form a maximal block of SNPs where there is almost no evidence of recombination: $|D'| = 1$ for all pairs of SNPs in the data after removing 2 of 120 YRI and 1 of 120 CEU haplotypes (the chimpanzee genome was used to define the ancestral alleles). A unique 'gene tree' was then built, and we used *genetree*[25], which assumes a coalescent prior on genealogies, to approximately infer ages for these mutations conditional on the data (a caveat is that the tree building does not account for the HapMap SNP ascertainment scheme). Because *genetree* assumes a randomly mating population, and the YRI represent almost all the HapMap haplotype diversity in this region, we ran the software (2,000,000 importance samples, otherwise default parameters) on the YRI data only and used this to construct Fig. 2c. Each node of the tree corresponds to a unique haplotype at these 19 SNPs, whose frequency in both CEU and YRI is shown at the base of the figure.

Motif searching. We tested all candidate motifs of 5 to 9 base pairs for enrichment in our African-enriched hotspot set relative to our shared hotspot set. We counted occurrences of all tested motifs in repeat and non-repeat backgrounds separately, and computed a separate *P* value for each genomic background with a chi-squared test, based on a contingency table that compares the counts of a particular motif to the counts of all motifs of that size. We converted each *P* value to a *Z* score, added the scores on each background, and then obtained a corresponding combined *P* value. Motifs were considered statistically significant only if they passed four stringent criteria: (1) they were statistically significant after Bonferroni correction for the number of motifs tested; (2) they were overrepresented in the African-enriched set; (3) they were statistically significant on both the repeat and non-repeat backgrounds ($P < 0.01$) independently; and (4) they were statistically significant when the joint *P* value was calculated only by comparing the frequency of the motif to other motifs of identical G/C content (to eliminate false positives due to any difference in G/C content between the hotspot sets). This testing revealed a unique significant motif, the 9-nucleotide oligomer CCCCAGTGA. We explored whether flanking DNA around exact matches to this motif also had a role by testing whether bases at a given site relative to the motif were associated with the difference in rates between African- and European-ancestry populations (Kruskal–Wallis test). Rates were evaluated in the 2 kb surrounding each motif occurrence. We separately evaluated flanking sequence using both the difference between YRI/CEU map rates, and the difference between the AE/S map rates, leading to the identification of the 17-bp consensus African-enriched motif (Supplementary Note 6 has full details). To identify close matches to this 17-bp motif among all matches to the 9-bp motif in the genome, for every occurrence of the 9-bp motif, we scored the flanking sequence bases proportionately to the relative increase in average crossover rate difference associated with each base, then multiplied across bases in the 17-mer region to provide an overall score. We ranked occurrences according to this score, and plotted rates around the top 500 (Fig. 3b). We verified these findings by measuring average crossover differences for each base using only odd chromosomes and used these to score motif occurrences on the (non-overlapping) set of even chromosomes, and vice versa (Supplementary Fig. 8).

***PRDM9* ZF length typing and genotyping of rs6889665.** To determine the number of ZF motifs of *PRDM9* in a subset of the samples used to build the map, published primer pairs[4] were used to amplify this region (forward: 5′-GGCCAGAAAGTGAATCCAGG-3′, reverse: 5′-GGGGAATATAAGGGGTCAGC-3′). Product lengths ranged between 7 and 20 repeats (801–1,893 bp). Four of the 166 African-American samples did not show an amplification product, presumably because of insufficient DNA quality. We also genotyped 90 YRI and 90 CEU HapMap samples.

The SNP rs6889665 was genotyped in the same samples using an allelic discrimination assay (forward primer: 5′-aaacttggaacatccatagggt-3′, reverse primer: 5′-cgaaaggagaaaagcataatcc-3′, Locked Nucleic Acid (LNA) probe 'C': 5′-/6-FAM/aGGGatAaatgaag/BHQ/-3′, LNA-probe 'T': 5′-/HEX/ AGAGatAaatGaagg/BHQ/-3′; LNA bases are given in capital letters). Reporter dyes: 6-FAM, 6-carboxyfluorescein; HEX, hexachlorofluorescein. Quencher: BHQ, Black Hole Quencher 1. Only one out of the 166 African-American samples failed in this assay. The same YRI and CEU samples as above were also genotyped.

26. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
27. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).
28. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).
29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).

Int. J. Cancer: **122,** 2581–2585 (2008)

Saturated fat intake predicts biochemical failure after prostatectomy

Sara S. Strom[1]*, Yuko Yamamura[1], Michele R. Forman[1], Curtis A. Pettaway[2], Stephanie L. Barrera[1] and John DiGiovanni[3]

[1]*Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, TX*
[2]*Department of Urology, The University of Texas M. D. Anderson Cancer Center, Houston, TX*
[3]*Department of Carcinogenesis, The University of Texas M. D. Anderson Cancer Center, Houston, TX*

Previous reports show that obesity predicts biochemical failure after treatment for localized prostate cancer. Since obesity is associated with increased fat consumption, we investigated the role that dietary fat intake plays in modulating obesity-related risk of biochemical failure. We evaluated the association between saturated fat intake and biochemical failure among 390 men from a previously described prostatectomy cohort. Participants completed a food frequency questionnaire collecting nutrient information for the year prior to diagnosis. Because fat and energy intake are highly correlated, the residual method was used to adjust fat (total and saturated) intakes for energy. Biochemical-failure-free-survival rates were calculated using the Kaplan–Meier method. Crude and adjusted effects were estimated using Cox proportional hazards models. During a mean follow-up of 70.6 months, 78 men experienced biochemical failure. Men who consumed high-saturated fat (HSF) diets were more likely to experience biochemical failure ($p = 0.006$) and had significantly shorter biochemical-failure-free-survival than men with low saturated fat (LSF) diets (26.6 *vs.* 44.7 months, respectively, $p = 0.002$). After adjusting for obesity and clinical variables, HSF-diet patients were almost twice as likely to experience biochemical failure (hazard ratio = 1.95, $p = 0.008$) compared to LSF diet patients. Men who were both obese and consumed HSF diets had the shortest biochemical-failure-free-survival (19 months), and nonobese men who consumed LSF diets had the longest biochemical-failure-free-survival (46 months, $p < 0.001$). Understanding the interplay between modifiable factors, such as diet and obesity, and disease characteristics may lead to the development of behavioral and/or targeted interventions for patients at increased risk of progression.
© *2008 Wiley-Liss, Inc.*

Key words: fat; diet; nutrition; prostate cancer; risk factor

The identification of modifiable factors that may influence long-term outcome for prostate cancer (PCa) has considerable potential to reduce morbidity and mortality.[1–3] Our group and others have reported that obesity is associated with increased risk of biochemical failure after treatment with radical prostatectomy[4,5] or external beam radiation[6] for localized disease. Since the prevalence of obesity in U.S. adults has reached epidemic proportions, furthering our understanding of the relationship between obesity-related risk and PCa outcome has become an increasingly important public health issue.

The epidemiological associations between high-fat diets and obesity,[7–9] and higher fat consumption with increased PCa risk and mortality[1,10–12] have been well documented. It has been suggested that some types of fat (*i.e.*, monounsaturated) may actually protect against PCa,[13–15] whereas saturated fat consumption has been more consistently associated with PCa risk, especially advanced disease.[11,16,17] To evaluate the role that dietary fat intake plays in modulating obesity-related PCa progression, we examined the association between self-reported dietary intake of saturated fat and biochemical failure in a well-defined cohort of PCa patients treated by radical prostatectomy.[5]

Subjects and methods

The study population is a subset of a previously described cohort of 526 patients at The University of Texas M.D. Anderson Cancer Center.[5] All patients had clinically organ-confined PCa at time of diagnosis and were treated with only prostatectomy. Due to the limited number of African–American and Hispanic participants, as well as known racial/ethnic variation in diet, we restricted the patient population to Caucasians ($N = 405$). This study was conducted in accordance with the Institutional Review Board, and informed consent was obtained prior to personal interview.

Using standardized questionnaires, demographic information, personal medical history, family history of cancer and other risk factor data were collected as previously described.[5] The semi-quantitative validated Block food frequency questionnaire (FFQ) (Health Habits and History Questionnaire), modified to incorporate foods commonly consumed in the Southwestern diet, was used to collect usual dietary intake for the year prior to diagnosis.[18] Patients were asked to report the average frequency of intake (per day, week, month or year) and usual portion size (*i.e.*, small, medium or large, relative to a defined medium portion) for ~180 food items. Approximately 80% of patients had the FFQ administered within 6 months of registration at M.D. Anderson. We did conduct a subset analysis and found no differences in range of responses between those who completed the FFQ within 6 months and those who completed it later. All patients were instructed by trained interviewers to provide answers for usual diet for the year prior to diagnosis. FFQs were reviewed by registered dietitians for completeness and acceptability. Only patients who completed the risk factor questionnaire and reported daily caloric intake between 600 and 5,000 kcal/day were included in this study ($N = 390$). DIETSYS+Plus (Version 5.9) along with the USDA National Nutrient Database for Standard Reference (Release 17) was used to calculate average daily intake of macro-nutrients and micro-nutrients for each individual.

Body mass index (BMI, kg/m^2) was calculated from self-reported height and weight. Obesity was defined according to the National Heart, Lung and Blood Institute guideline of BMI ≥ 30.0 kg/m^2. Leisure time physical activity was categorized based on participant response to "the year before your diagnosis, how often did you do physical activities such as jogging, biking or brisk walking (long enough to get sweaty)?" Family history of PCa in first-degree relatives was defined as PCa diagnosed in father, brother or son.

Clinico-pathologic characteristics were abstracted by trained study personnel from medical records using standardized forms and included prostatectomy Gleason score, pathological stage (including surgical margin status and seminal vesicle involvement) and preoperative PSA levels.[5] Tumors were classified based on pathological stage as pT2 (organ-confined) and pT3 (extraprostatic extension +/− seminal vesicle invasion). Time to progression was measured from date of prostatectomy to date of 1st detectable prostate specific antigen (PSA) test (≥0.1 ng/ml, biochemical failure) or last date the patient was known to have no evidence of disease (censor).

Grant sponsor: National Cancer Institute; Grant numbers: CA84964, CA90270; Grant sponsor: National Institute of Environmental Health Sciences; Grant number: EHS07784; Grant sponsor: Department of Defense; Grant number: DAMD 17-98-1-84711.

***Correspondence to:** Department of Epidemiology, Unit 1340, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. Fax: +713-792-9568.
E-mail: sstrom@mdanderson.org

Received 1 August 2007; Accepted after revision 23 November 2007
DOI 10.1002/ijc.23414
Published online 6 March 2008 in Wiley InterScience (www.interscience.wiley.com).

Statistical analysis

Categorical variables, such as family history, history of diabetes, leisure-time physical activity, prostatectomy Gleason score, margin status and pathological stage were analyzed using χ^2 or Fisher's exact tests to evaluate differences in the distribution of the clinical, demographic and risk factor data. Continuous variables, such as age, BMI, education and dietary intake, were compared between groups using Student's *t*-tests. Since total and saturated fat intake were highly correlated ($r = 0.89$ and $r = 0.86$, respectively, $p < 0.001$ for both) with total daily energy intake, we energy-adjusted fat consumption using the residual method.[19] Energy-adjusted total and saturated fat intakes were categorized into quartiles for initial analyses. Since risk of progression was significantly higher among men in the upper quartiles of total and saturated fat consumption (*i.e.*, Q4) as compared to those in the lower 3 quartiles (*i.e.*, Q1–Q3), analyses were conducted by dichotomizing intake(high intake = Q4, lower intake = Q1–Q3). To assess whether total fat or saturated fat was a better predictor of outcome, parallel predictive models were constructed. Total energy intake (kcal) was also evaluated as an independent predictor of outcome modeled as a continuous variable. The Gleason scores from prostatectomies were analyzed in 4 categories: 6, 7 (3 + 4), 7 (4 + 3) and ≥8. The distribution of preoperative PSA was skewed to the right, therefore, all values were log-transformed prior to analysis, and PSA was analyzed as a continuous variable.

Biochemical-failure-free survival rates were calculated using the Kaplan–Meier method, and log-rank tests were used to evaluate statistical significance. Univariate Cox proportional hazards models allowed us to evaluate the crude effects of each factor of interest. Variables with $p \leq 0.10$ were evaluated for inclusion in a multivariable model that simultaneously adjusted for all other included variables. In a forward stepwise manner, the multivariable model was constructed; and 95% confidence intervals were estimated for all point estimates using 2-sided testing (SPSS version 12.0, Chicago, IL). The final multivariable model only includes factors shown to significantly improve the predictive value.

Results

This subset of 390 men was representative of the previously described larger cohort with respect to age and clinico-pathologic characteristics.[5] Table I shows patient characteristics for men by level of saturated fat intake [high saturated fat diets (HSF) and lower in saturated fat (LSF)]. Compared to men who consumed LSF diets, men who consumed HSF diets were younger (59.4- *vs.* 61.2-years-old, respectively; $p = 0.03$) and had higher BMIs at diagnosis (28.4 *vs.* 27.3 kg/m^2, respectively; $p = 0.03$) (Table I). There were no statistically significant differences in clinico-pathologic characteristics (*i.e.*, prostatectomy Gleason score, PSA, or pathological stage), family history of PCa, education, history of diabetes or physical activity between these 2 groups. As expected, men consuming HSF diets, also consumed more calories (2,292 *vs.* 2,088 kcal/day, respectively, $p = 0.04$) and total fat (102 *vs.* 73 g/day, respectively, $p < 0.001$) compared to men who ate LSF diets (Table I). The top contributors to daily intake of saturated fat for this patient population were beef steaks, cheese and cheese spreads, hamburgers and cheeseburgers, eggs, ice cream and salad dressing/mayonnaise.

During the follow-up period (mean = 97.3 months), 20% of the patients with pathologically organ-confined disease experienced biochemical failure. Biochemical failure-free survival was estimated using Kaplan–Meier survival methods stratified by saturated fat intake (Fig. 1*a*). Men who ate HSF diets were significantly more likely to experience biochemical failure ($p = 0.006$), and had significantly shorter biochemical failure-free survival than men who consumed less saturated fat (26.6 *vs.* 44.7 months, respectively, $p = 0.004$). Five years after surgery, about 65% of men who consumed HSF diets had no evidence of disease compared to 80% of men who consumed LSF diets. Initial analyses of the risk of progression indicated that men in the 2nd and 3rd quartiles of energy-adjusted total and saturated fat intake had no appreciable change in risk compared to the lowest quartile. For this reason, fat intake (both total and saturated fat) were dichotomized as Q4 *vs.* Q1–Q3.

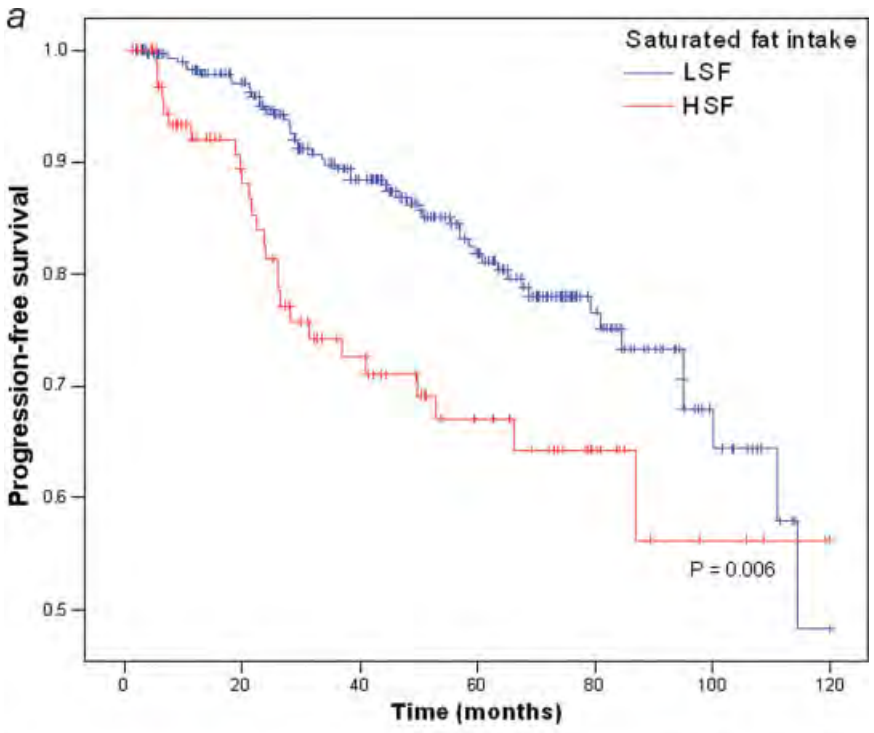
TABLE I – PARTICIPANT CHARACTERISTICS

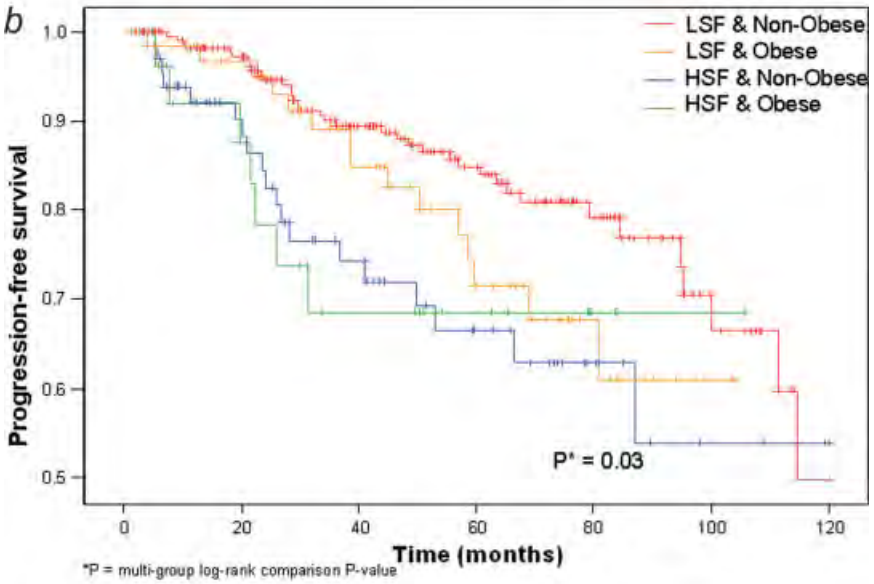
Variable	Low saturated fat ($N = 293$)	High saturated fat ($N = 97$)	*p*-value
Age (mean ± SD)	61.2 ± 6.8	59.4 ± 7.3	0.03
Education (years, mean)	15.3	15.4	0.89
+ Family history of PCa in FDR[1]	58 (19.8)	26 (26.8)	0.15
BMI at D_x (kg/m^2, mean)	27.3	28.4	0.03
Diabetes diagnosis	11 (4.0)	8 (8.4)	0.09
Leisure time physical activity			
1+ times/wk	218 (74.4)	65 (67.7)	
Few times/m	33 (11.3)	10 (10.4)	
Rarely/Never	42 (14.3)	21 (21.9)	0.22
Gleason score			
6	73 (24.9)	30 (30.9)	
7 (3 + 4)	90 (30.7)	20 (20.6)	
7 (4 + 3)	64 (21.8)	23 (23.7)	
8	66 (22.5)	24 (24.7)	0.27
PSA > 10 ng/ml	56 (19.6)	20 (21.7)	0.65
+ Surgical margin	41 (14.2)	18 (18.8)	0.28
pT3/T4	76 (26.1)	30 (31.3)	0.33
Calories (kcal/day)	2087.9	2292.1	0.04
Fat (g/day)	73.1	101.6	<0.001
Saturated fat (g/day)	23.4	37.2	<0.001
Unsaturated fat (g/day)	49.6	64.4	<0.001
% Energy fat	31.0	39.6	<0.001
% Energy saturated fat	9.9	14.5	<0.001
PCa progression (%)	17.7	26.8	0.05

[1]FDR, first-degree relatives.

Using Kaplan–Meier methods, we evaluated the combined effects of obesity and saturated fat consumption (Fig. 1*b*). Men who were both obese and consumed HSF diets had the shortest biochemical failure-free survival (19 months), and nonobese men who consumed LSF diets had the longest biochemical failure-free survival (46 months; $p < 0.001$). Nonobese men who ate HSF diets and obese men who ate LSF diets had intermediate progression-free survival times (29.4 and 41.5 months, respectively). Approximately 85% of nonobese men on LSF diets were biochemical failure-free at 5 years after surgery, compared to 70% obese on LSF and about 65% of nonobese and obese men on HSF diets. The interaction between saturated fat intake and obesity was not statistically significant ($p = 0.99$).

We used Cox proportional hazards models to simultaneously adjust for relevant clinico-pathologic variables in a multivariable Cox proportionate hazards model (Table II). We found that energy-adjusted HSF diet remained an independent predictor of biochemical failure in our final model; PCa patients who consumed HSF diets were almost twice as likely to experience biochemical failure compared to men who ate less saturated fat (HR = 1.98, $p = 0.006$). Increased BMI (continuous) was modestly associated with increased risk of BF (HR = 1.05, $p = 0.05$). Since lack of physical activity may be associated with increased BMI and consuming poorer diet (*i.e.*, diet high in saturated fat), we evaluated the predictive utility of including leisure-time physical activity in the multivariable model; however, physical activity did not improve the overall fit of the model and was not included in the final model. Multivariable analyses indicated that inclusion of energy-adjusted saturated fat intake explained a greater proportion of variance as indicated by the log likelihood of that model compared to the model including total energy intake. Saturated and total fat intake were significantly correlated ($r = 0.95$, $p < 0.001$). However, saturated fat intake explained significantly more overall variance in the model compared to total fat. The addition of total fat intake into the multivariable model with saturated fat intake had no appreciable impact on the overall goodness of fit of the model and was not included.





c

Saturated Fat	BMI	Mean time to progression (months)
Low	Non-obese	45.9
Low	Obese	41.5
High	Non-obese	29.4
High	Obese	19.2

FIGURE 1 – (*a*) Progression-free survival by saturated fat intake (low vs. high) (LSF = low saturate fat intake; HSF = high saturated fat intake), (*b*) Progression-free survival by saturated fat intake and BMI (Obese = BMI > 30 kg/m^2; Non-obese = BMI < 30 kg/m^2), (*c*) Mean time to progression in months by saturated fat intake and BMI.

In initial multivariable models, energy intake alone was evaluated as a potential predictor of failure. Parallel models incorporating the same covariates and either total energy intake or energy-adjusted saturated fat intake were constructed and compared; the model with energy-adjusted saturated fat explained more variance in the data and was better at predicting outcome compared to the one with total energy intake. In contingency table analysis, no association was found between energy-adjusted saturated fat intake and total energy intake. The inclusion of energy in the multivariable model neither significantly improved

TABLE II – MULTIVARIABLE MODEL OF BIOCHEMICAL FAILURES

Variable	Hazard ratio	95% CI
Without energy		
High saturated fat intake	1.95[1]	1.19–3.19
BMI (kg/m^2, continuous)	1.06[1]	1.00–1.12
With energy		
High saturated fat intake	1.90[2]	1.16–3.11
BMI (kg/m^2, continuous)	1.06[2]	1.00–1.12

[1]Adjusted for pathologic stage, surgical margin involvement and Gleason score.–[2]Adjusted for pathologic stage, surgical margin involvement, Gleason score and total energy intake.

the fit of the model nor affected the point estimates (Table II); therefore energy intake was removed from the final model.

Discussion

Our results showed that high prediagnostic saturated fat intake was associated with a 2-fold increased risk of biochemical failure in this cohort of 390 Caucasian men with localized PCa treated with prostatectomy. The multivariable model indicated that this increase in risk of biochemical failure was independent of the increased risk associated with obesity, and both obese and nonobese men who consumed HSF diet had shorter biochemical failure-free survival.

Some epidemiological studies found a direct association between saturated fat intake and PCa risk and prognosis, especially in advanced disease,[16] suggesting that saturated fat may play a role in PCa prognosis. However, not all studies have adjusted for the effects of total energy intake, and the associations or lack thereof reported in these studies may be partially attributable to residual confounding. Additionally, our data support the findings reported by Meyer *et al*. that prediagnostic HSF intake was associated with increased PCa mortality.[11] However, to our knowledge, no studies have evaluated the combined effects of both energy-adjusted saturated fat intake and obesity as predictors of PCa progression.

The mechanisms by which these associations affect PCa prognosis have not been established, although some studies suggest that alterations in insulin metabolism may be involved.[20] In overweight and obese nondiabetic men, diets high in saturated fat were shown to induce insulin resistance, which has been suggested to play a role in prognosis.[21] Additionally, it has been shown that men, whose diets were highest in saturated fat had the highest levels of IGF-1 and lowest levels of IGFBP-3 compared to men who ate diets lower in saturated fat.[22] Castrated xenograft mice injected with LAPC-4, an androgen-sensitive PCa cell line, and fed an isocaloric low-fat diet had significantly lower serum levels of insulin and IGFBP-1/-2 as well as slower PCa progression compared to similarly treated mice on high-fat diet.[20]

Another plausible mechanism by which saturated fat may influence PCa progression involves heterocyclic amine consumption since several key contributors to saturated fat intake (*i.e*., beef steaks and hamburgers/cheeseburgers) are known to have high levels of heterocyclic amines. These foods are often prepared using high-heat generating methods, such as grilling or broiling, which has been shown to significantly increase dietary intake of heterocyclic amines, particularly, 2-amino-1-methyl-6-phenylimidazo[4,5-b]-pyridine (PhIP), previously demonstrated to have carcinogenic properties. Human prostate tissue is capable of activating heterocyclic amines that can then bind to DNA and form adducts, which have been associated with prostate carcinogenesis.[23] Additionally, PhIP-DNA adducts levels, a quantitative measurement of PhIP exposure, have been demonstrated to show an association with greater tumor volume and higher Gleason score among African–Americans,[24] both of which have been shown to be associated with PCa progression. Higher PhIP intake has been significantly associated with increased PSA levels, which is also a predictor of PCa outcome.[25] We were unable to evaluate heterocyclic amine consumption since information on cooking methods was not collected; however, future studies are being designed to collect and incorporate these data.

Sex hormone levels have been shown to be influenced by saturated fat intake. Dietary intervention studies in healthy men have shown that a low-fat diet decreased androgen levels both in serum and urine[26] and a high fat diet increased plasma and urinary testosterone and DHEA-S.[27] These results demonstrate the ability of short-term changes in fat intake to directly affect the hormonal milieu known to play a key role in the natural history of PCa.[28] Overall, the evidence suggests that saturated fat might affect PCa prognosis through several inter-related mechanisms and other dietary components may act in concert or discordance.

This study has some limitations. Nutritional data were collected at the time of study enrollment, and we do not have quantifiable information about how patients changed their diets since diagnosis. There is potential for measurement error since the FFQ is semiquantitative; however, this error should be minimized as we used the data from the FFQs simply to categorize men as high or low consumers of nutrients rather than compare absolute values. Our patient population was limited to Caucasians, as we did not have sufficient power to evaluate inter-racial/ethnic variation in dietary intake in conjunction with progression *vs*. no-progression.

On the other hand, this study has several strengths. The patients comprising our cohort were all diagnosed with clinically localized disease, received the same treatment and did not have adjuvant therapy postoperatively prior to biochemical failure. Since all participants in this study are cancer patients interviewed at baseline (*i.e*., prior to biochemical failure), there should be no difference in recall between patients who experienced biochemical failure and those who did not. Restricting our patient population to Caucasians limits the effects of inter-racial/ethnic variation in food consumption patterns as well as other lifestyle and genetic differences that may help reduce the effects of confounding.

These results expand upon our previous finding that obesity was associated with increased risk of biochemical failure following prostatectomy, and suggest that saturated fat intake plays a role in PCa progression. After duplicating these findings in a larger patient population from different racial/ethnic groups, future interventions may be designed to decrease consumption of dietary saturated fat to reduce risk of progression in PCa patients as has been done for breast cancer patients.[29] It is our hope that these results can be integrated into clinical practice to identify patients at high-risk of progression following definitive therapy. Increasing our understanding of the interplay between modifiable factors, such as lifestyle (*e.g*., diet) and disease characteristics, may lead to developing targeted interventions for patients at increased risk for biochemical failure.

Acknowledgements

The authors thank the patients who participated in this research and the physicians in the Department of Urology for allowing us to interview their patients.

References

1. Colli JL, Colli A. Comparisons of prostate cancer mortality rates with dietary practices in the United States. Urol Oncol 2005;23:390–8.
2. Giovannucci E, Liu Y, Platz EA, Stampfer MJ, Willett WC. Risk factors for prostate cancer incidence and progression in the health professionals follow-up study. Int J Cancer 2007;121:1571–8.

3. Chan JM, Holick CN, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, Giovannucci EL. Diet after diagnosis and the risk of prostate cancer progression, recurrence, and death (United States). Cancer Causes Control 2006;17:199–208.
4. Freedland SJ, Grubb KA, Yiu SK, Humphreys EB, Nielsen ME, Mangold LA, Isaacs WB, Epstein JI, Partin AW. Obesity and risk of biochemical progression following radical prostatectomy at a tertiary care referral center. J Urol 2005;174:919–22.
5. Strom SS, Wang X, Pettaway CA, Logothetis CJ, Yamamura Y, Do KA, Babaian RJ, Troncoso P. Obesity, weight gain, and risk of biochemical failure among prostate cancer patients following prostatectomy. Clin Cancer Res 2005;11:6889–94.
6. Strom SS, Kamat AM, Gruschkus SK, Gu Y, Wen S, Cheung MR, Pisters LL, Lee AK, Rosser CJ, Kuban DA. Influence of obesity on biochemical and clinical failure after external-beam radiotherapy for localized prostate cancer. Cancer 2006;107:631–9.
7. Munoz de Chavez M, Chavez A. Diet that prevents cancer: recommendations from the American Institute for Cancer Research. Int J Cancer Suppl 1998;11:85–9.
8. Kennedy ET, Bowman SA, Powell R. Dietary fat intake in the US population. J Am Coll Nutr 1999;18:207–12.
9. Satia-Abouta J, Patterson RE, Schiller RN, Kristal AR. Energy from fat is associated with obesity in U.S. men: results from the Prostate Cancer Prevention Trial. Prev Med 2002;34:493–501.
10. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer 1975;15:617–31.
11. Meyer F, Bairati I, Shadmani R, Fradet Y, Moore L. Dietary fat and prostate cancer survival. Cancer Causes Control 1999;10:245–51.
12. Kim DJ, Gallagher RP, Hislop TG, Holowaty EJ, Howe GR, Jain M, McLaughlin JR, Teh CZ, Rohan TE. Premorbid diet in relation to survival from prostate cancer (Canada). Cancer Causes Control 2000;11:65–77.
13. Norrish AE, Jackson RT, Sharpe SJ, Skeaff CM. Men who consume vegetable oils rich in monounsaturated fat: their dietary patterns and risk of prostate cancer (New Zealand). Cancer Causes Control 2000;11:609–15.
14. Kushi L, Giovannucci E. Dietary fat and cancer. Am J Med 2002;133:63S–70S.
15. Ramon JM, Bou R, Romea S, Alkiza ME, Jacas M, Ribes J, Oromi J. Dietary fat intake and prostate cancer risk: a case-control study in Spain. Cancer Causes Control 2000;11:677–8.
16. Kristal AR, Cohen JH, Qu P, Stanford JL. Associations of energy, fat, calcium, and vitamin D with prostate cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11:719–23.
17. Whittemore AS, Kolonel LN, Wu AH, John EM, Gallagher RP, Howe GR, Burch JD, Hankin J, Dreon DM, West DW, Teh C-Z, Paffenbarger RS, Jr. Prostate cancer in relation to diet, physical activity, and body size in Blacks, Whites, and Asians in the United States and Canada. J Natl Cancer Inst 1995;87:652–61.
18. Strom SS, Yamamura Y, Duphorne CM, Spitz MR, Babaian RJ, Pillow PC, Husting SD. Phytoestrogen intake and prostate cancer: a case-control study using a new database. Nutr Cancer 1999;33:20–5.
19. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:17–27.
20. Ngo TH, Barnard RJ, Anton T, Tran C, Elashoff D, Heber D, Freedland SJ, Aronson WJ. Effect of isocaloric low-fat diet on prostate cancer xenograft progression to androgen independence. Cancer Res 2004;64:1252–4.
21. Xiao C, Giacca A, Carpentier A, Lewis GF. Differential effects of monounsaturated, polyunsaturated and saturated fat ingestion on glucose-stimulated secretion, sensitivity and clearance in overweight and obese, non-diabetic humans. Diabetologia 2006;49:1371–9.
22. Gunnell D, Oliver SE, Peters TJ, Donovan JL, Persad R, Maynard M, Gillatt D, Pearce A, Hamdy FC, Neal DE, Holly JM. Are diet-prostate cancer associations mediated by the IGF axis? A cross-sectional analysis of diet, IGF-1 and IGFBP-3 in healthy middle-aged men. Br J Cancer 2003;88:1682–6.
23. Williams JA, Martin FL, Muir GH, Hewer A, Grover PL, Phillips DH. Metabolic activation of carcinogens and expression of various cytochromes P450 in human prostate tissues. Carcinogenesis 2000;21:1683–9.
24. Tang D, Liu JJ, Bock CH, Neslund-Dudas C, Rundle A, Savera AT, Yang JJ, Nock NL, Rybicki BA. Racial differences in clinical and pathological associations with PhIP-DNA adducts in prostate. Int J Cancer 2007;121:1319–24.
25. Bogen KT, Keating GA, II, Chan JM, Paine LJ, Simms EL, Nelson DO, Holly EA. Highly elevated PSA and dietary PhIP intake in a prospective clinic-based study among African-Americans. Prostate Cancer Prostatic Dis 2007;10:261–9.
26. Wang C, Catlin DH, Starcevic B, Heber D, Ambler C, Berman N, Lucas G, Leung A, Schramm K, Lee PW, Hull L, Swerdloff RS. Low-fat high-fiber diet decreased serum and urine androgens in men. J Clin Endocrinol Metab 2005;90:3550–9.
27. Dorgan JF, Judd JT, Longcope C, Brown C, Schatzkin A, Clevidence BA, Campbell WS, Nair PP, Franz C, Kahle L, Taylor PR. Effects of dietary fat and fiber on plasma and urine androgens and estrogens in men: a controlled feeding study. Am J Clin Nutr 1996;64:850–5.
28. Chan JM, Gann PH, Giovannucci EL. Role of diet in prostate cancer development and progression. J Clin Oncol 2005;23:8152–60.
29. Chlebowski RT, Blackburn GL, Thomson CA, Nixon DW, Shapiro A, Hoy MK, Goodman MT, Giuliano AE, Karanja N, McAndrew P, Hudis C, Butler J, et al. Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. J Natl Cancer Inst 2006;98:1753–5.